Contrast-reversed binocular dot-pairs in random-dot stereograms for depth perception in central visual field: Probing the dynamics of feedforward-feedback processes in visual inference

Li Zhaoping

University of Tübingen, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Published 2021 in *Vision Research*, Vol. 186, Pages 124-139.

Abstract: In a random-dot stereogram (RDS), the spatial disparities between the interocularly corresponding black and white random dots determine the depths of object surfaces. If a black dot in one monocular image corresponds to a white dot in the other, disparity-tuned neurons in primary visual cortex (V1) respond as if their preferred disparities become non-preferred and vice versa, reversing the disparity sign reported to higher visual areas. Reversed depth is perceptible in the peripheral but not the central visual field. This study demonstrates that, in central vision, adding contrast-reversed dots to a noisy RDS (containing the normal contrast-matched dots) can augment or degrade depth perception. Augmentation occurs when the reversed depth signals are congruent with the normal depth signals to report the same disparity sign, and occurs regardless of the viewing duration. Degradation occurs when the reversed and normal depth signals are incongruent with each other *and* when the RDS is viewed briefly. These phenomena reflect the Feedforward-Feedback-Verify-and-reWeight (FFVW) process for visual inference in central vision, and are consistent with the central-peripheral dichotomy that central vision has a stronger top-down feedback from higher to lower brain areas to disambiguate noisy and ambiguous inputs from V1. When a RDS is viewed too briefly for feedback, augmentation and degradation work by adding the reversed depth signals from contrast-reversed dots to the feedforward, normal, depth signals. With a sufficiently long viewing duration, the feedback vetoes incongruent reversed depth signals and amends or completes the imperfect, but congruent, reversed depth signals by analysis-by-synthesis computation.

Keywords: Central vision, contrast-reversed random-dot stereogram, central-peripheral dichotomy (CPD), reversed-depth perception, top-down feedback, primary visual cortex (V1), analysis-by-synthesis, Feedforward-Feedback-Verify-and-reWeight (FFVW).

1 Introduction

Julesz (1971) popularized the use of the random-dot stereograms (RDSs) for studying stereo vision. In a RDS, depth cues are unavailable monocularly, and are only available in the correspondence between dots in different eyes. The spatial disparity between the corresponding dots, one in the left eye and the other in the right eye, gives the depth signal. Fig. 1A shows a schematic of a RDS containing a central disk of dots in front of a surrounding ring of dots in a three-dimensional (3D) scene. Each dot in the ring occupies the same location in the two monocular images that are viewed by the two eyes; this defines these dots as having a zero binocular disparity. For each dot in the central disk, the location of its image in the left eye is displaced relative to that in the right eye. This displacement is called binocular disparity and is indicated in Fig. 1A by an arrow pointing from the right-eye dot to the corresponding left-eye dot. The rightward pointing arrow defines a positive disparity, consistent with the disk being in front of the surrounding ring in 3D space, i.e., being nearer to the viewer than the surrounding ring. A negative disparity would be represented by a leftward pointing arrow, and would characterize a disk that lay behind the surrounding ring.

Neurons in the primary visual cortex (V1) are tuned to binocular disparities in visual input stimuli such as gratings, bars, or RDSs (Ohzawa et al., 1990; Qian, 1994; Cumming and Parker, 1997). Each disparity-tuned neuron has a disparity tuning curve — its response is higher to its preferred disparities than to its non-preferred disparities. For example, some neurons prefer positive disparity and thus respond more to nearer depth surfaces, whereas other neurons prefer negative disparity and respond more to farther surfaces. Patterns of such neural responses send information about surface depths from V1 to higher brain areas. Conventional RDSs involve dots whose contrast polarities are matched between the two eyes (as in Fig. 1A). In this paper, we call the depth signals emanating from V1 based on such contrast-matched RDSs normal depth signals.

However, V1 neurons also respond to contrast-reversed inputs, including RDSs, when the corresponding inputs in the two eyes have opposite contrast polarity, such that, e.g., a black dot in one eye corresponds to a white dot in the other eye(Ohzawa et al., 1990; Cumming and Parker, 1997). The neural responses to contrast-reversed RDSs are lower to the disparity that is usually preferred, and higher to the disparity that is usually not preferred. In other words, consider a neuron that is excited by the positive disparity of a nearer surface in a contrast-matched RDS like Fig. 1A (and suppressed by a negative disparity for a farther-surface). This neuron will then be suppressed by the same positive disparity (and excited by the same negative disparity) if the contrast-matched dots are replaced by contrast-reversed dots without changing the disparity. Hence, to a contrast-reversed RDS, neural responses signal reversed depth to higher brain areas. In this paper, we call these V1 responses reversed depth signals.

For many years, it has been known that humans cannot perceive the reversed depth reported by V1 in response to the contrast-reversed RDSs(Cumming et al., 1998; Read and Eagle, 2000; Doi et al., 2011; Asher and Hibbard, 2018). For example, if each contrast-matched dot-pair for the central disk in Fig. 1A is replaced by a contrast-reversed dot-pair at the same respective monocular image location (and thus keeping the disparity unchanged), human observers would not be able to tell whether the central disk is in front of, or behind, the surrounding ring (when the RDS is viewed in the central visual field). Here, we show that the contrast-reversed dot-pairs can still impact depth perception when they are mixed with contrast-matching dot-pairs. Furthermore, we use these observations in various visual input situations to probe the feedforward

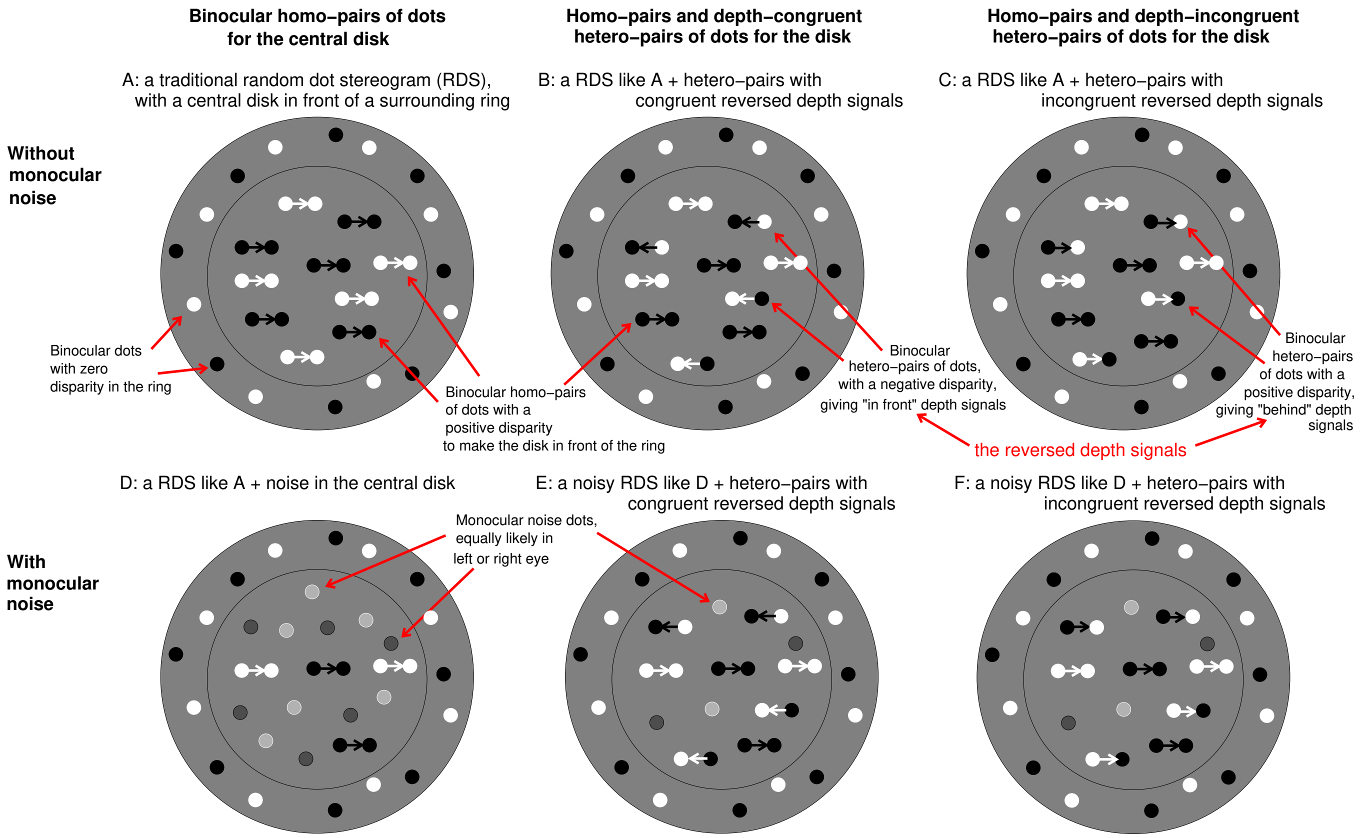
Binocular homo−pairs of dots
for the central disk
Homo−pairs and depth−congruent
hetero−pairs of dots for the disk
Homo−pairs and depth−incongruent
hetero−pairs of dots for the disk
A: a traditional random dot stereogram (RDS),
with a central disk in front of a surrounding ring
B: a RDS like A + hetero−pairs with
congruent reversed depth signals
C: a RDS like A + hetero−pairs with
incongruent reversed depth signals
Without
monocular
noise
Binocular dots
with zero
disparity in the ring
Binocular homo−pairs
of dots with a
positive disparity
to make the disk in front of the ring
Binocular
hetero−pairs of dots,
with a negative disparity,
giving "in front" depth signals
Binocular
hetero−pairs
of dots with a
positive disparity,
giving "behind" depth
signals
the reversed depth signals
D: a RDS like A + noise in the central disk
E: a noisy RDS like D + hetero−pairs with
congruent reversed depth signals
F: a noisy RDS like D + hetero−pairs with
incongruent reversed depth signals
Monocular noise dots,
equally likely in
left or right eye
With
monocular
noise

Figure 1: Schematics of six kinds of RDSs. (A): A traditional RDS, depicting a central disk surface of dots in front of the surrounding ring of dots. The concentric circular contours are for illustration only and are not part of the visual input. Black and white dots in the ring depict binocular dots having zero disparity. In the central disk, a binocular dot with a positive disparity is represented using a pair of contrast-matched dots linked by a rightward arrow pointing from the right-eye dot to the corresponding left-eye dot. Each such a pair of dots is called a homo-pair since the two dots in a pair have the same contrast polarity (black or white). (B) like A, but the disk also contains some contrast-polarity-reversed pairs, or hetero-pairs, of dots (a white dot in one eye and a black dot in the other eye). The two dots in a hetero-pair are linked by an arrow pointing leftward to depict a negative binocular disparity. However, the hetero-pairs signal, via V1 responses, a reversed depth, i.e., for a disparity in the opposite direction to that of the arrow. Thus, both homo-pairs and hetero-pairs in (B) indicate, in their associated V1 responses, that the disk is in front of the ring. (C) Like B, except that the hetero-pairs have a positive disparity, so that the reversed depth signals indicate that the disk is behind the ring, incongruent with the normal depth signals from the homo-pairs. D, E, F): like A, B, C, respectively, except that the central disk contains also monocular noise dots. Each noise dot is randomly shown to either the left or right eye, and is randomly black or white (visualized by a black or white circle with a darker or lighter shade inside, respectively). Noise dots do not correspond with any dot in the other eye, and are not due to depth occlusion. In a dense RDS, each noise dot could accidentally match up with other stimulus dots (including the binocularly corresponding dots) in the other eye to generate a perception of random and noisy depth dots in a 3D scene.

and feedback interactions between V1 and higher visual areas.

Before we proceed further, it is useful to define some terms. In the literature, contrast-matched RDSs are often called correlated RDSs and contrast-reversed RDSs are often called anti-correlated RDSs. However, if a RDS contains a mixture of contrast-matched and contrast-reversed pairs of binocularly corresponding dots, the overall binocular correlation could be positive or negative depending on whether the contrast-matched pairs dominate. So to avoid confusion, we will not use the term correlated or anti-correlated to describe any RDS. A binocularly corresponding pair of dots will be explicitly called a contrast-polarity-matched pair or a contrast-polarity-reversed pair. To shorten the term, it is also called a contrast-matched pair or a contrast-reversed pair, or simply a homo-pair or a hetero-pair, respectively. A RDS (or a surface patch in a RDS) will be described by two quantities, one is f_{homo}, the fraction of dots that belong to homo-pairs, and the other is f_{hetero}, the fraction of dots that belong to hetero-pairs. A RDS with $f_{homo} > 0$ and $f_{hetero} = 0$ will be called a contrast-matched RDS; while a RDS with $f_{homo} = 0$ and $f_{hetero} > 0$ will be called a contrast-reversed RDS.

The fact that human observers typically cannot see reversed depth in a contrast-reversed RDS is consistent with the idea that V1 is not a site of consciousness(Crick and Koch, 1995). However, Zhaoping and Ackermann (2018) showed that the reversed depth can be perceived in the peripheral visual field, as predicted by the central-peripheral dichotomy (CPD) which was originally proposed on the basis of computational and psychophysical arguments(Zhaoping, 2017). CPD considers the top-down feedback from higher to lower visual areas (such as V1) to aid visual recognition, particularly in challenging situations such as noisy, ambiguous, or partially occluded visual inputs. CPD suggests that this feedback is stronger in central vision and weaker or absent in peripheral vision. According to this proposal, perceptual inference in central vision is a form of analysis-by-synthesis(Helmholtz, 1925; MacKay, 1956; Yuille and Kersten, 2006). V1 responses are feedforward signals reporting, e.g., an input binocular disparity to suggest to higher visual areas initial hypotheses for a perceptual decision about, e.g., the depth of an underlying object surface. If sensory inputs are unclear or ambiguous in central vision, multiple hypotheses that are in conflict with each other can be suggested and given substantial weights in the feedforward signal. Each of the initial hypotheses is re-evaluated by the brain in three steps. First, according to brain's internal model or prior knowledge of the visual world, higher brain centers synthesize a would-be visual input that should resemble the actual visual input if the perceptual hypothesis suggested by the feedforward signals is correct. Second, the synthesized input is fed back to lower visual areas such as V1 for comparison with the actual visual input. Third, the strength of the initial perceptual hypothesis is reweighted according to the degree of the match between the synthesized and the actual inputs, such that a good or poor match strengthens or weakens, respectively, the initial hypothesis for the ultimate perceptual outcome. These steps are called the Feedforward-Feedback-Verify-reWeight (FFVW) process(Carpenter and Grossberg, 1987; Zhaoping, 2017). Critically, the verification occurs in lower visual areas via feedback since, due to an attentional bottleneck starting at V1's output according to a recent proposal(Zhaoping, 2019), not all visual input information available in V1 is sent forward to higher visual areas. Because of this bottleneck, the feedforward signals from V1 are even more ambiguous than the retinal signals about the properties of the visual scene.

Consider the application of FFVW to a visual input containing a contrast-reversed RDS with a positive disparity for a near depth surface. V1 neurons respond as if the input disparity is negative, and feed an initial hypothesis for a far surface to higher areas. These higher areas synthesize inputs as being from homo-pairs of dots on a far surface according to their internal model of the world. The synthesized inputs are fed back to V1, and they fail to match the actual inputs containing the hetero-pairs. Consequently, the initial hypothesis of a far depth, the reversed depth, fails the verification and is therefore weakened or vetoed by the reweighting, making it hard for observers to perceive a far depth surface. According to the CPD, the feedback in FFVW is weaker in the peripheral visual field. Consequently, peripheral reversed depth signals in the feedforward inputs are less likely to be vetoed, making the reversed depth more likely perceived. Accordingly, peripheral vision is more vulnerable to visual illusions suggested by misleading inputs from V1(Zhaoping, 2019). Indeed, illusions analogous to the reversed depth, such as reversed phi motion(Anstis, 1970) and the flip tilt illusion(Zhaoping, 2020), caused by V1 neural responses to hetero-pairs of stimulus correspondence in two different time points or spatial locations, are stronger or only visible in peripheral vision.

In this paper, we infer the feedforward and feedback dynamics of FFVW in central vision by examining how reversed depth signals from hetero-pairs of dots influence perception of the depth generated by homo-pairs of dots. Although the reversed depth of a surface generated purely by hetero-pairs of dots is typically invisible in central vision, the feedforward signals that they generated could combine with those of homo-pairs if the latter are also present. For example, if the reversed depth signals from the hetero-pairs agree with the normal depth signals from the homo-pairs (this occurs when the disparity in hetero-pairs is the negative of the disparity in the homo-pairs), as in Fig. 1BE, the combined feedforward depth signals from V1 could be stronger. We can then expect, for example, that depth perception will be stronger or clearer for such a RDS compared to another RDS that has the same homo-pairs of dots without the hetero-pairs of dots. Conversely, if the reversed depth signals from the hetero-pairs are opposite to the normal depth signals from the homo-pairs (when the hetero- and homo-pairs of dots have the same disparity, see Fig. 1CF), we could expect depth perception to be weaker or less clear.

Explicitly, we consider that each patch of RDS, e.g., the RDS patch for the central disk in Fig. 1, can have up to three kinds of image dots: (1) dots from homo-pairs which create conventional normal depth signals, (2) dots from hetero-pairs which create reversed depth signals, and (3) monocular noise dots, which can be in the left-eye or right-eye image, that do not correspond with any dot in the monocular image of the other eye (for simplicity, we ignore monocular dots due to occlusion here). In each monocular image, we use f_{homo}, f_{hetero}, and f_{noise} to denote the fractions of dots arising from homo-pairs, hetero-pairs, and noise, respectively, so that

$$f_{homo} + f_{hetero} + f_{noise} = 1. \tag{1}$$

For simplicity, and unless stated explicitly otherwise, in the current study, we restrict the disparity in the hetero-pairs of dots to be uniformly either the negative of, or the same as, the disparity of the homo-pairs. Correspondingly, the reversed depth signal in the hetero-pairs is, respectively, in agreement with (when homo- and hetero-pairs have the opposite disparity), or the opposite of (when homo- and hetero-pairs have the same disparity), the depth signal associated with the homo-pairs. The corresponding RDS will be called depth-congruent or depth-incongruent, or incongruent or congruent for short. A RDS without any hetero-pairs will be called depth-neutral or neutral for short. To illustrate, the characteristics and notations for the central disk in each RDS in Fig. 1 are listed in Table 1.

In general, this study uses RDSs containing noise dots such as those shown in Fig. 1DEF which are neutral, congruent, and incongruent, respectively. Unless stated explicitly, all the RDSs used in the study

Table 1: Characteristics and notations for the random-dot stereograms (RDSs) in Fig. 1

	Fractions f_{homo}, f_{hetero}, and f_{noise} of dots			disparities $d_{\rm homo}$ and $d_{\rm hetero}$
	for depth	for reversed-depth	for noise	in homo- and hetero-pairs
Noiseless RDSs				
Fig. 1A, a neutral RDS,	$f_{homo} = 1$,	$f_{hetero} = 0$,	$f_{noise} = 0$	no hetero-pairs
Fig. 1B, a congruent RDS,	$f_{homo} < 1$,	$f_{hetero} > 0$,	$f_{noise} = 0$	$d_{\rm homo} = -d_{\rm hetero}$
Fig. 1C, an incongruent RDS,	$f_{homo} < 1$,	$f_{hetero} > 0$,	$f_{noise} = 0$	$d_{\rm homo} = d_{\rm hetero}$
Noisy RDSs				
Fig. 1D, a neutral RDS,	$f_{homo} < 1$,	$f_{hetero} = 0$,	$f_{noise} > 0$	no hetero-pairs
Fig. 1E, a congruent RDS,	$f_{homo} < 1$,	$f_{hetero} > 0$,	$f_{noise} > 0$	$d_{\rm homo} = -d_{\rm hetero}$
Fig. 1F, an incongruent RDS,	$f_{homo} < 1$,	$f_{hetero} > 0$,	$f_{noise} > 0$	$d_{\rm homo} = d_{\rm hetero}$

have the same net density of stimulus dots when we include all dots regardless of whether they arise from homo-pairs, hetero-pairs, or noise dots. Viewing such a noisy RDS, observers typically find the depth surface of the central disk noisy. A noise dot in one eye can accidentally match a noise or even a non-noise dot in the other eye, generating ghost depth dots that can be perceived when the viewing duration is not too brief. See Fig. 2 for examples. As we will see later, two neutral RDSs (with $f_{hetero} = 0$) with sufficiently different signal-to-noise ratios f_{homo}/f_{noise} are also perceived as being differentially noisy. However, given a noisy RDS, even non-naive observers find it hard to tell without scrutiny whether the noisy RDS contains hetero-pairs of dots. Hence, consider two noisy RDSs which have the same set of homo-pairs of dots, e.g., both of them have $f_{homo} = 50\%$. One RDS is neutral (hence its $f_{noise} = 50\%$), and the other is congruent (or incongruent) with (e.g.,) $f_{noise} = 25\%$ and $f_{hetero} = 25\%$. We can ask which of these two RDSs yields a stronger or clearer percept of the depth surface. If the hetero-pairs of dots behave like noise dots for depth perception, then the two RDSs should result in similar quality of depth percept, thus appearing equally noisy. Otherwise, one RDS should appear less noisy than the other. For example, the congruent RDS may appear less noisy, and/or subjects may find it more difficult to discriminate depth in the incongruent RDS.

Hence, to answer our scientific question of whether the reversed depth signals produced by the hetero-pairs have any perceptual impact that is different from the impact of noise dots, we compare depth perception for neutral, congruent, and incongruent noisy RDSs that have the same f_{homo} and the same density of total stimulus dots. Such noisy RDSs that differ in their composition and nature of noise dots and hetero-pairs of dots, are illustrated schematically in Fig. 1DEF, and shown in experimental form in Figs. 2ABC and Figs. 2DEF.

Fig. 3A depicts plausible hypotheses about the perceptual effects of the reversed depth: (1) “no effect”, depicted by the blue line, when the depth percepts are equally clear in the neutral, congruent, and incongruent RDSs having the same f_{homo}, (2) “pro-effect”, when the depth percept is clearer in the congruent RDS than the neutral RDS, and (3) “anti-effect”, when the depth percept is less clear in the incongruent RDS than the neutral RDS. One plausible hypothesis is that pro- and anti-effects are simultaneously present. The “no effect” hypothesis is expected from the invisibility of the reversed depth in a pure contrast-reversed RDS in central vision, and suggests an overwhelming feedback veto on the feedforward reversed depth signals from V1. The other two effects suggest that the feedback is ineffective, or does not completely veto the reversed depth signals, or utilizes the reversed depth signals in perceptual inferences in a way aligned with

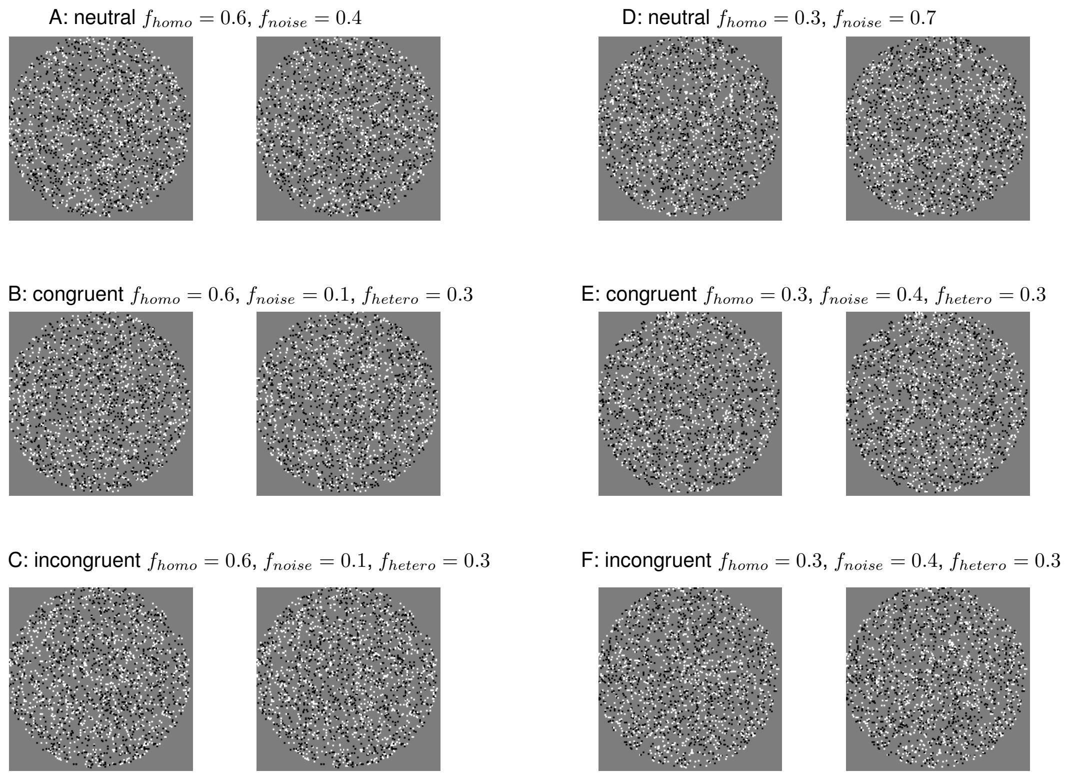
A: neutral $f_{homo} = 0.6$, $f_{noise} = 0.4$
D: neutral $f_{homo} = 0.3$, $f_{noise} = 0.7$
B: congruent $f_{homo} = 0.6$, $f_{noise} = 0.1$, $f_{hetero} = 0.3$
E: congruent $f_{homo} = 0.3$, $f_{noise} = 0.4$, $f_{hetero} = 0.3$
C: incongruent $f_{homo} = 0.6$, $f_{noise} = 0.1$, $f_{hetero} = 0.3$
F: incongruent $f_{homo} = 0.3$, $f_{noise} = 0.4$, $f_{hetero} = 0.3$

Figure 2: Six example RDSs of the type in the experiments using a smaller size for the stimulus dots. Each RDS, for a central disk in front of the surrounding ring in 3D, contains a pair of images displayed side by side, the left image shown to the left eye and the right image to the right eye. Each RDS is characterized by the fractions f_{homo}, f_{hetero}, and f_{noise}, respectively, of stimulus dots for the central disk that arise from homo-pairs of binocular (contrast-matched) dots, hetero-pairs of binocular (contrast-reversed) dots, and random monocular noise dots that do not arise from occlusion or correspondence between the two eyes, and $f_{homo} + f_{hetero} + f_{noise} = 1$ by definition. (The surrounding ring has only homo-pairs of binocular dots as in all the experimental stimuli.) The right three RDSs are noisier (with a smaller f_{homo} and a larger f_{noise}) than the left ones. Free fusing the left and right images in each RDS, readers can experience that the effect of the reversed depth signals by the hetero pairs depends on the signal level f_{homo} relative to the f_{homo} needed for each observer to perceive the depth of the central disk.

the feedforward signals.

Fig. 3BC illustrate schematically the predictions the three hypotheses make about the behavioral outcomes from Experiments 1 and 2 in this study. Experiment 1 asks observers to report for which of two RDSs is the percept of the depth surface of the central disk clearer or less noisy. The critical pairs of RDSs differ monotonically in congruency (but not in f_{homo}) so that one is neutral and the other is incongruent (neutral-incongruent, or N-I for short), or one is congruent and the other is neutral (congruent-neutral, or C-N for short), or one is congruent and the other is incongruent (congruent-incongruent, or C-I for short). According to the “no effect” hypothesis, the probability that observers report the more congruent RDS as clearer should be 50%. According to the other hypotheses, this probability should be larger than 50%. Experiment 1 also included control trials with two neutral RDSs (neutral-neutral, or N-N for short), for which the one with

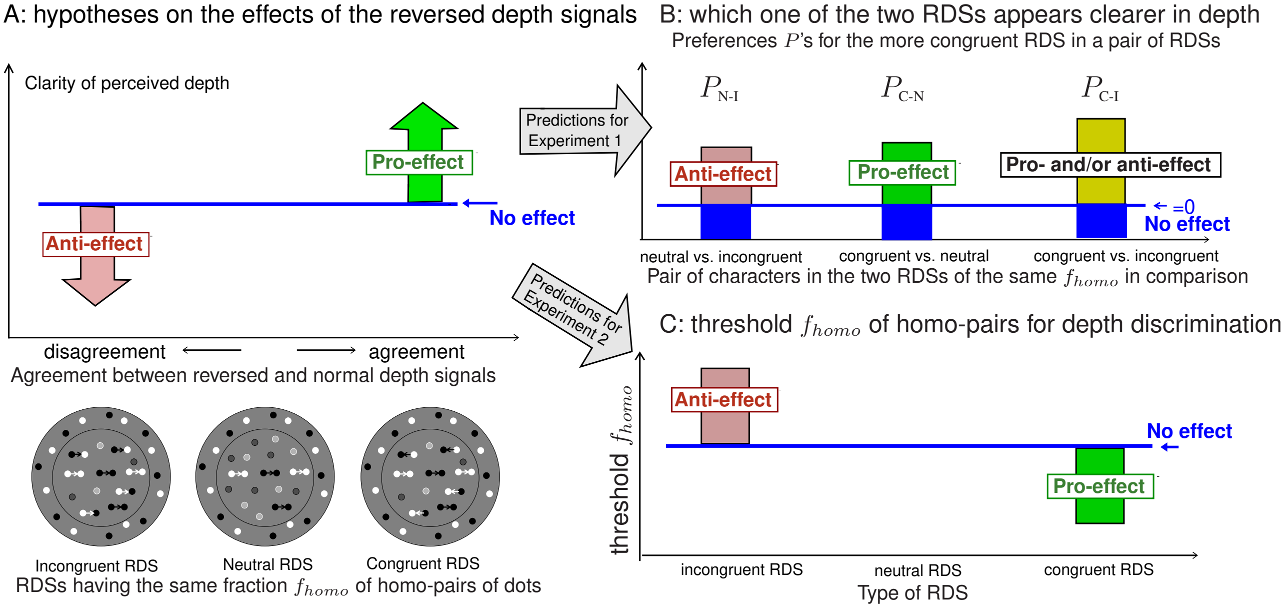
A: hypotheses on the effects of the reversed depth signals
Clarity of perceived depth
Pro-effect
No effect
Anti-effect
disagreement
agreement
Agreement between reversed and normal depth signals
Incongruent RDS
Neutral RDS
Congruent RDS
RDSs having the same fraction f_{homo} of homo-pairs of dots
Predictions for Experiment 1
B: which one of the two RDSs appears clearer in depth
Preferences P's for the more congruent RDS in a pair of RDSs
$P_{\text{N-I}}$
$P_{\text{C-N}}$
$P_{\text{C-I}}$
Anti-effect
Pro-effect
Pro- and/or anti-effect
=0
No effect
neutral vs. incongruent
congruent vs. neutral
congruent vs. incongruent
Pair of characters in the two RDSs of the same f_{homo} in comparison
Predictions for Experiment 2
C: threshold f_{homo} of homo-pairs for depth discrimination
threshold f_{homo}
Anti-effect
No effect
Pro-effect
incongruent RDS
neutral RDS
congruent RDS
Type of RDS

Figure 3: Plausible hypotheses concerning the perceptual effects of the reversed depth signals and their respective predictions of experimental outcomes. A: plausible hypotheses for the clarity of perceived depth versus the type (incongruent, neutral, or congruent) of RDS given a fixed fraction f_{homo} of homo-pairs of dots. The "no effect" hypothesis (horizontal blue line) suggests that the clarity of the perceived depth is the same across the three types of RDSs. The "pro-effect" is that the congruent RDS will appear clearer in depth than the neutral RDS. The "anti-effect" is that the incongruent RDS will appear less clear in depth than the neutral RDS. Four plausible hypotheses could include "no effect", "pro-effect" only, "anti-effect" only, and "pro- and anti-effects". B & C: predictions of the hypotheses in A by the correspondingly colored bars. B: two RDSs of the same f_{homo} (and same f_{hetero} if both non-neutral) are compared for clarity in perceived depth in Experiment 1; the preference P (defined in equations (2-6)) for the more congruent RDS being clearer is equal to, or greater than, zero, depending on the hypothesis in A. $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$ quantifies the preference P for the more congruent RDS in a neutral-incongruent, congruent-neutral, or congruent-incongruent pair of RDSs, respectively. $P_{\text{N-I}} > 0$ when the anti-effect is present, $P_{\text{C-N}} > 0$ when the pro-effect is present, $P_{\text{C-I}} > 0$ when at least one of the pro- and anti-effects is present. C: the threshold or minimum fraction (f_{homo}) of homo-pairs of dots needed to discriminate the depth order in a noisy RDS is measured in Experiment 2. This threshold should be lower for a congruent than a neutral RDS if the pro-effect is present, and be higher for an incongruent than a neutral RDS if the anti-effect is present. Both Experiment 1 and 2 probed whether and how the measured quantities depend on experimental manipulations that make top-down feedback more or less effective.

fewer noise dots should be clearer.

To be explicit, we define a signal preference P for the more congruent RDS (or the less noisy member of a neutral-neutral pair) as follows.

In a pair of RDSs in comparison,
one RDS is reported as clearer in perceived depth,
let $F_1 \equiv$ fraction of trials reporting
the more congruent or less noisy RDS,
$F_2 \equiv$ fraction of trials reporting the
less congruent or noisier RDS $= 1 - F_1$,
signal preference $P \equiv F_1 - F_2$, often specifically denoted (2)
as $P_{\text{N-N}}$ when the RDS pair is neutral-neutral, (3)
as $P_{\text{N-I}}$ when the RDS pair is neutral-incongruent, (4)
as $P_{\text{C-N}}$ when the RDS pair is congruent-neutral, and (5)
as $P_{\text{C-I}}$ when the RDS pair is congruent-incongruent. (6)

We call these P's signal preferences, referring to the depth signals sent by V1 neurons to higher visual areas. Hence, P is the degree to which perception prefers a RDS with a stronger depth signal relative to another RDS with a weaker depth signal. When the two RDSs have the same f_{homo} value but differ in reversed depth signals, we write the signal preferences as $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$, and the preference is due to different f_{hetero} values and their different congruencies. When the two RDSs are both neutral, the signal preference is written as $P_{\text{N-N}}$ and comes from the different normal depth signals associated with the different f_{homo} values in the two RDSs. $P_{\text{N-N}}$ assesses the discrimination of the clarity of perceived depth in RDSs without reversed depth signals.

In Experiment 2, we vary the fraction f_{homo} of homo-pair dots in a RDS to find the threshold level of f_{homo} needed to accurately identify the depth order defined by the disparity in the homo-pairs (see Fig. 3C). The "no effect" hypothesis predicts that this threshold does not depend on the RDS's congruency. The pro- and anti-effects predict that the threshold will be lower in a congruent RDS and higher in an incongruent RDS, respectively, compared to the threshold in a neutral RDS.

We found that the reversed depth signals from hetero-pairs do indeed impact the perceived depth according to a pro-effect, and in some situations also according to an anti-effect. To examine whether these effects change in a way consistent with FFVW, we employed two methods for weakening the anticipated feedback. One is to shorten the viewing duration of a static RDS (in Experiment 1), or to shorten the duration of each frame in a dynamic RDS containing multiple RDS image frames (in Experiment 2; each frame contains an independently generated random set of stimulus dots, keeping constant the overall disparity and densities of various dots). Since it takes time for the feedback to occur and interact with the feedforward signals, shorter viewing durations could impact the magnitude of the effects. Another method is to make the stimulus dots smaller, assuming that it is more difficult for the feedback process to verify whether the actual input stimulus matches the top-down expected would-be input stimulus when the dots are smaller.

The next sections report the details of the experimental methods and findings; the last section presents a summary and discussion.

2 Experimental Methods

In this paper, each RDS always contained a central disk and a surrounding ring. All the experimental trials had a radius $r = 3.61^o$ for the disk, an outer radius $R = 4.7^o$ for the ring, a zero disparity for the ring, a disparity difference 0.087^o between the ring and the disk, and a fraction $f = 25\%$ of image area (for the ring or disk) that would be covered by the stimulus dots if they did not occlude each other. The RDS's statistical characteristics that varied between intervals of stimulus presentations within a trial, between trials, or between experimental sessions include (1) whether the RDS is neutral, congruent, or incongruent; (2) its f_{homo} and f_{hetero} values (while $f_{noise} = 1 - f_{homo} - f_{hetero}$ by definition) characterized by a two component vector (f_{homo}, f_{hetero}); (3) the viewing duration T; (4) whether the disk was in front or behind; (5) whether the square-shaped dots in a RDS is larger (with a side length of 0.174^o) or smaller (with a side length of 0.087^o); and (6) whether the RDS is static (when a single pair of dichoptic image was viewed for the whole duration T) or dynamic (when the set of stimulus dots for the RDS was replaced every 0.02 seconds by another random and independent set of stimulus dots while keeping all the statistical properties of the RDS unchanged).

This study is approved by the Ethics Council of the Max Planck Society and the Ethik-Kommission an der Medizinischen Fakultät der Eberhard-Karls-Universität und am Universitätsklinikum Tübingen.

2.1 Experiment 1: design and the psychophysical task

Human observers were asked to view two static RDS stimuli (like those in Fig. 2) in each trial. The two RDSs differed from each other in f_{homo}, or in f_{hetero}, or in whether it is congruent or incongruent. The depth orders of the disk in the two RDSs were randomly and independently generated. The two RDSs were presented in two time intervals, each interval had a duration T which was fixed in each experimental session, and the two intervals were separated by a gap of 1 second. For each trial, randomly either one of the two RDSs was chosen to be presented in the first interval and the other RDS in the second interval. The random set of stimulus dots in each interval was independently generated from the random set in the other interval or those of any other trials.

The observers were asked to take their time and to give two reports for each trial by pressing buttons after the second interval. First, they had to report whether the RDS in the first or second interval looked relatively clearer in terms of the depth surface and the depth order of the central disk relative to the surrounding ring. Second, they had to report, for the RDS that they had just reported as clearer, whether the central disk was in front or behind the surrounding ring. Observers could freely move their gaze over the stimuli.

A RDS in each interval of each trial can be neutral (when $f_{hetero} = 0$), congruent, or incongruent. The non-neutral RDSs always had $f_{hetero} = 0.2$ in Experiment 1 reported in this paper. The pair of the RDS types in each trial can be neutral-neutral (control), neutral-incongruent, congruent-neutral, or congruent-incongruent. Each neutral-neutral pair is characterized by N-N$(f_{homo}, \Delta f_{homo})$, in which f_{homo} is that of the RDS with a smaller f_{homo} and Δf_{homo} is the difference between the f_{homo} values in the two RDSs. In each non-control pair, the two RDSs share the same f_{homo} value, and when one RDS is congruent and the other is incongruent then they also share the same f_{hetero} value. Hence a non-control pair is characterized as N-I(f_{homo}, f_{hetero}), C-N(f_{homo}, f_{hetero}), or C-I(f_{homo}, f_{hetero}), for a neutral-incongruent, congruent-neutral, or congruent-incongruent pair, respectively. A set of the three different kinds of non-control pairs, N-I, C-N, and C-I, that share the same (f_{homo}, f_{hetero}) parameters is called a triplet of (non-control) pairs. Each experimental session randomly interleaved trials from seven differently characterized RDS pairs, one control pair and two triplets (differing in f_{homo}) of non-control pairs. With $f_{hetero} = 0.2$ fixed for the non-control pairs, then the set of RDSs in each session can be characterized by the parameter set (f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), f_{homo}(triplet 2)), with "control", "triplet 1", "triplet 2" of the parameters referring to whether it is for the control pair or for one of the triplets. Each session had 364 trials in a random order, 52 trials for each of the seven differently characterized RDS pairs, and each observer completed a session in 5 blocks with rests between the blocks.

Before the start of each session, the observer was given one or more dozens of practice trials using only neutral-neutral pairs of RDSs. The experimenter (the author) adjusted the $(f_{homo}, \Delta f_{homo})$ values for every dozen of these practice trials to assess the sensitivities of the observer to f_{homo} and Δf_{homo} and to enable the observer to experience various degrees of the task difficulty. The parameters f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), and f_{homo}(triplet 2) for the testing trials in each session were set according to the requirements of the experimental design, or according to the experimenter's estimate of the observer's ability based on these practice trials and this observer's performance in any previous sessions of Experiment 1.

The parameters $\vec{f} \equiv (f_{homo}(\text{control}), \Delta f_{homo}(\text{control}), f_{homo}(\text{triplet } 1), f_{homo}(\text{triplet } 2))$ for data sessions contributing to Figs. 4 – 6 are listed in Table 2, in which each observer is denoted by a unique symbol used to plot his/her individual data points in the figures.

Table 2: Stimulus parameters $\vec{f}$ for data sessions used in Figs. 4 – 6

$\vec{f}$	observer	figure(s) or figure part(s)
(0.30, 0.30, 0.60, 0.40)	△	Fig. 4
(0.55, 0.20, 0.75, 0.65)	△	Figs. 4 – 6
(0.35, 0.25, 0.60, 0.45)	×	Figs. 5 – 6
(0.40, 0.20, 0.60, 0.45)	+	Figs. 5 – 6
(0.25, 0.25, 0.50, 0.25)	◊	Figs. 5 – 6
(0.60, 0.20, 0.80, 0.75)	◁	Fig. 5, Fig. 6A bottom
(0.40, 0.20, 0.60, 0.45)	◁	Fig.5A for T=1 second
(0.25, 0.20, 0.45, 0.30)	□	Figs.5 – 6
(0.20, 0.20, 0.40, 0.20)	□	Fig.5A for T=0.02 second

2.2 Stimuli for Experiment 1

The arrangement of the equipment was identical to that in the previous papers(Zhaoping and Ackermann, 2018; Zhaoping, 2012, 2017), except that eye tracking was not used. Equipment, including a Mitsubishi 21-inch cathode-ray tube (CRT), and a mirror stereoscope, were purchased from Cambridge Research System (CRS), and were calibrated (e.g., gamma correction of the CRT's signal-to-luminance relationship) by CRS software. The viewing distance was 50 centimeters.

The RDSs were made using the same method as in (Zhaoping and Ackermann, 2018). The monocular images to the two eyes were displayed side by side on the CRT and viewed by the respective eyes via a mirror stereoscope as described in (Zhaoping, 2012). The gray background, white dots, and black dots on the CRT had luminance values around 50, 100, and $0\ \mathrm{cd/m}^2$, respectively. Vergence was anchored by a black rectangular frame in each monocular image enclosing an area 17.8^o in width and 15^o in height, with the frame thickness of 0.22^o. The monocular image for the RDS was centered at the center of the respective rectangular frame for each eye. Each dot in the RDS was a square with a side length 0.174^o in most experimental sessions, and a side length 0.087^o in some sessions.

To make a stimulus, we start from making a RDS without any hetero-pairs or monocular noise dots using the previous method(Zhaoping and Ackermann, 2018). First, we started with two concentric disks (with radius r for the disk and radius $R > r$ for the ring) of random dots. For each disk, each dot was placed at any location within the disk with equal probability and was set to be black or white with equal probability. A random sequential order is assigned to these dots of each disk, so that each dot, after being assigned to be a homo-dot, a hetero-dot, or a noise dot in the procedure specified below, is drawn into the two monocular images by this same sequential order, possibly occluding any previously drawn dots that are sufficiently close. The two monocular images of the RDS were made from these two disks of dots as follows. In the unit of pixel size, the disparity of the central disk was an integer $d = 2$ pixels. The left and right monocular images contained the dots from the smaller disk (radius r) after these dots were shifted horizontally by 1 and -1 pixel, respectively. For each monocular image, the dots for the surrounding ring were those from the larger disk, excluding any dot that was either within the image area of the shifted smaller disk or would overlap in the monocular image with any dots from the smaller disk. This results in a RDS for which the central disk has only homo-pairs and is in front of the ring. If the disk should be behind the ring, the left and right images were then swapped.

To make monocular noise dots, a fraction f_{noise} of the original homo-pairs for the central disk were selected at random. For each of these selected pairs, the binocular correspondence between the two dots in the pair was removed by independently at random reassigning the locations of the two resulting monocular dots, one for the left eye and the other for the right eye, in a central disk area of the same radius as the central disk and concentric with the surrounding ring. To make the hetero-pairs, another fraction f_{hetero} of the original homo-pairs was randomly selected; in each of the selected pair of two dots, randomly one of the dots (i.e., from the left or right eye at random) was assigned to the opposite contrast polarity (from white to black or black to white). If the RDS should be congruent, then the image locations of the two dots in the two monocular images were swapped to switch the sign of the disparity between the two dots.

Each test trial started with a binocular (zero-disparity) text "press any button to start the next trial " displayed at the center of the frame that anchored vergence (this anchoring frame was present throughout an experimental session). The following sequence of stimuli followed after the button press: (1) a blank screen other than the vergence anchoring frame in the gray background for 0.7 second; (2) the RDS for the first stimulus interval for a duration T; (3) a blank screen as in (1) for one second; (4) the RDS for the second stimulus interval for a duration T; (5) a blank screen like that in (1) with two binocular text strings "First clear" and "Second clear" displayed near the left and right border of the vergence anchoring frame (not overlapping with the display location where the RDS had been) to indicate to the observer to press the left or right button to choose if the first or the second RDS appeared clearer for the perceived depth of the central disk. After the observer pressed a left or right button, these two strings were replaced by binocular strings “Front disk” and “Back disk” respectively at the respective locations to prompt the observer to press the left or right button to indicate whether the disk in their chosen RDS was in front or behind the ring.

2.3 Experiment 2: design, the psychophysical task, and the stimuli

In Experiment 2, the observers were shown a RDS containing the same ring and central disk (along with the same, ever present, vergence anchoring frame) as in Experiment 1. Each trial contained only one interval showing a RDS for a duration T = 0.2 second. The observer’s task was to take their time to report, after the RDS disappeared, whether the central disk was in front or behind the ring by pressing one of the two buttons (one closer to them and one further away). As in Experiment 1, each trial started with a binocular text string “press any button to start the next trial”, this string disappeared upon the observer’s button press to start a trial, and the RDS appeared one second later.

In each session, the RDS in each trial was randomly one of the six (2×3) types, it could be neutral, congruent, or incongruent, and for each case it could be static or dynamic. In the static case, the RDS contained a single dichoptic image pair shown for the whole duration T =0.2 second. In the dynamic case, the random set of stimulus dots in the RDS was independently regenerated every 0.02 second while the other stimulus characters of the RDS stimulus were fixed. Each pair of dichoptic images was generated in the same way as that in Experiment 1. In each trial, the central disk was randomly in front or behind the ring.

For each non-neutral RDS, $f_{hetero} = 0.3$ was used. Each session started with a dozen or more practice trials on neutral RDSs that had a large enough f_{homo} value so that the observers could perform almost all the practice trials correctly. During the test trials, the values f_{homo} for the six RDS types were adjusted independently in parallel using staircase methods to obtain the threshold f_{homo} needed to see depth clearly. A 7-down-1-up staircase was used so that the task was not too difficult for the observers. The staircase adjustment was carried out in each session as follows. Let $f_{homo}(i)$ denote the f_{homo} value for the RDS type i, with $i = 1, 2, ..., 6$, this $f_{homo}(i)$ value was the same across i for the first trial of each i in the session. Let the current trial about to be executed to be of RDS type i. If the last trial of RDS type i had an incorrect depth report, then $f_{homo}(i)$ was raised by the transform $f_{homo}(i) \rightarrow 1.1 f_{homo}(i)$ for the current trial; if the depth reports in the last seven trials of this RDS type i were all correct, then $f_{homo}(i)$ was lowered by the transform $f_{homo}(i) \rightarrow 0.9 f_{homo}(i)$ for the current trial; if neither of the previous two requirements was met, $f_{homo}(i)$ stayed unchanged. The constraint $f_{homo} + f_{hetero} + f_{noise} = 1$ was always maintained, so that every adjustment of the f_{homo} was accompanied by an adjustment of $f_{noise} = 1 - f_{homo} - f_{hetero}$ (and each fraction was non-negative and never more than 1). In very rare cases during the staircase (this occurred for one observer only and for only seven trials total) when $f_{homo} > 0.7$ for a non-neutral RDS, then $f_{hetero} = 1 - f_{homo}$ and $f_{noise} = 0$ were set for the trial, and such trials were not used for calculating the threshold fractions of homo-pairs needed for depth perception. Each observer participated in two sessions that had the same initial f_{homo}, and so performed a total of 300 trials for each RDS type. Each session comprised multiple blocks of about 10 minutes each with breaks between the blocks. The initial f_{homo} for the first session was estimated for each observer from his/her performance during the practice trials at the beginning of the session or from his/her performance in Experiment 1.

2.4 Data analysis method

In each session of Experiment 1, $N = 52$ trials were collected for each of the seven types of pairs of RDSs. Let n_1 be the number of trials an observer reported the less noisy or the more congruent RDS of the pair as being clearer in perceived depth. Let $n_2 = N - n_1$ be the number of trials in which the observer reported the noisier or less congruent RDS as being clearer. Then the signal preference P by equation (2) is determined by $F_1 = n_1/N$ and $F_2 = n_2/N$. For the results reported in Fig. 4, the error of this estimated P is calculated as $2\sqrt{F_1 F_2/N}$ based on binomial statistics. The probability p that n_1 is no less than its observed value if the observer randomly and with equal probability reported either of the two RDSs in a trial as clearer is calculated as $p = \sum_{n \geq n_1}^{N} 0.5^N N!/n!(N-n)!$, and the preference P is considered as significantly larger than zero (in Fig. 4) when $p < 0.05$.

In Experiment 1, the accuracy of the depth report for each RDS (with its particular fractions of various dots) in each RDS pair was the fraction of trials in which the depth report was correct, among the trials in which this particular RDS was reported as having a clearer perceived depth.

Fig. 5A and Fig. 6A show the preferences $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$ for each particular viewing duration T and stimulus dot size, for individual observers or for the average across observers. Each plotted $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$ for each observer is obtained after pooling all the trials of that particular congruency combination, N-I, C-N, or C-I, from two different triplets that differed in f_{homo} values in a single data session. In a couple of cases, a single observer had two sessions of data for a given viewing duration T and a given stimulus dot size (see the Result section for details). In this case, the values of $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$ plotted for this observer are the result of pooling trials from the corresponding congruency combination from four different triplets that differed in f_{homo} values. Fig. 5B.1 and Fig. 6B.1 plot $P_{\text{N-N}}$. Each bar in Fig. 5 and Fig. 6 is the average of the corresponding quantity (e.g., $P_{\text{N-N}}$) across the observers, and the error bar on each bar marks the standard error of this average. This observer-averaged quantity (e.g. $P_{\text{N-N}}$) is considered significantly larger than zero when a t-test on the set of these quantities (e.g., $P_{\text{N-N}}$), one per individual observer, has $p < 0.05$. Two observer-averaged quantities in each plot of Fig. 5 and Fig. 6 are considered significantly different from each other if a matched-sample t-test over the two corresponding lists of individual quantities has $p < 0.05$.

In Experiment 2, from all the trials for each observer and each of the six types of RDS, the following method is used to obtain the threshold f_{homo} needed to see the depth order clearly. Let f_1, f_2, f_3, ..., be the $f_{homo} \leq 1 - f_{hetero}$ values tested, each f_i (for $i = 1, 2, ...$) had n_i trials of which $m_i \leq n_i$ trials had correct depth order reports. The threshold, denoted as f_{th}, is then obtained by the following maximum-likelihood method. Let the response data be generated by an underlying psychometric function $p(f, f_{th}, b, \lambda)$ such that the probability of giving a correct response at $f_{homo} = f$ follows this Weibull function (with parameters f_{th}, b, and λ)

$$p(f, f_{th}, b, \lambda) = 0.5 + (0.5 - \lambda)\left[1 - e^{-\left(\frac{f}{f_{th}}\right)^b}\right]. \tag{7}$$

Let $p_i \equiv p(f = f_i, f_{th}, b, \lambda)$, then the probability of getting m_i correct responses out of n_i trials across various f_i is

$$\text{probability} = \prod_i \frac{n_i!}{m_i!(n_i - m_i)!} p_i^{m_i} (1 - p_i)^{n_i - m_i}. \quad (8)$$

The set of parameters f_{th}, b, and λ in $p(f, f_{th}, b, \lambda)$ that maximizes the logarithm of this probability is then obtained by an optimization procedure (using the *fmincon* function from Matlab). The resulting threshold f_{th} is reported in Fig. 7 for each observer and each RDS type. Each bar in Fig. 7 is the average of the corresponding values across observers, the error bar on each bar is the standard error of this average. Two such averages are said to be significantly different from each other, or different from a certain value, when $p < 0.05$ is obtained from a (matched-sample) t-test on the list(s) of individual observer values.

2.5 Observers

Altogether six adult observers aged between 26 and 56 participated in at least one experiment. Among them, two were males and one was the non-naive author (female) who also collected all the data. The observers had normal or corrected-to-normal vision, and could see depth in conventional RDSs. The experimental work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained for experimentation with each observer.

3 Results

3.1 Experiment 1: perceptual effects of reversed depth signals examined by the clarity of the perceived depth

Six observers participated in Experiment 1. In each trial, they compared the relative clarity of depth percepts in two RDSs. The two RDSs differed in the presence and/or congruency of reversed depth signals coming from hetero-pairs of dots, or, in control (neutral-neutral) trials, they differed in the fraction (f_{homo}) of homo-pairs of dots. The observers' preferences (defined by equations (2-6)) for the less noisy or more congruent RDS are reported in Figs. 4, 5, and 6. Data from different observers are represented by differently shaped data point symbols $+$, $\times$, ◇, △, □, ◁. Symbol □ marks data from the only non-naive observer, the author.

Each session randomly interleaved 7×52 trials from seven differently characterized RDS pairs, one neutral-neutral (control) pair and two triplets (that differed in f_{homo}) of pairs (neutral-incongruent, congruent-neutral, congruent-incongruent) in which the two RDSs differed in the presence or congruency of the reversed depth signals but had the same f_{homo}. Each experimental session used a particular viewing duration T, which could take the value of 0.02, 0.1, 0.2, or 1.0 second, and each session used a particular size of the stimulus dot, which was usually a $0.174^o \times 0.174^o$ square but a $0.087^o \times 0.087^o$ square in some sessions. Often, the same observer and the same set of RDS characters (in terms of f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), and f_{homo}(triplet 1), see Method) were involved in different sessions that differed in T or in the stimulus dot size in order to examine better whether and how the perceptual effects of the reversed depth signals change with the viewing durations T and the dot size.

3.1.1 Signal preferences due to reversed depth signals depend on the signal level f_{homo} or the accuracy of the depth order reports

Fig. 4 shows an example naive observer's performance in two sessions that differed in overall f_{homo} levels, with viewing duration T = 0.02 second. In neutral-neutral (control) pairs of RDSs, there was no reversed depth signals, and this observer could clearly discriminate between clarities of depth percepts caused by different amounts of signal f_{homo} (while $f_{noise} = 1 - f_{homo}$). This is shown in Fig. 4A for two neutral-neutral pairs, one for each session. In one pair, the two RDSs had $f_{homo} = 0.3$ and 0.6 respectively; in the other pair, they had $f_{homo} = 0.55$ and 0.75. For each pair, the preference $P_{\text{N-N}}$ (defined in equations (2-6)) for the less noisy RDS was significant and at least 0.35, implying that the less noisy RDS was reported in at least 2/3 of the trials as giving a clearer depth percept. The report on the depth order (obtained only for the RDS reported as clearer in a trial) was accurate for most trials, except for the noisiest (when $f_{homo} = 0.3$) of the four RDSs.

Meanwhile, when a trial compared two RDSs that differed in the presence or congruency of the reversed depth signals, shown in Fig. 4B, preferences P for the more congruent RDS were significant only when the depth reports were sufficiently accurate. Specifically, the preferences $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$ (for neutral-incongruent, congruent-neutral, congruent-incongruent RDS pairs), manifesting, respectively, the anti-effect, the pro-effect, and both the pro- and anti-effects, and their average $\bar{P}$ are plotted. Sufficiently accurate depth reports occurred, as expected, when f_{homo} in a RDS pair was large enough. (In control trials comparing two neutral RDSs, preference for the less noisy RDS was also insignificant if both RDSs evoked inaccurate depth reports in some pilot data sessions.) Each plot in Fig. 4B shows preferences associated with a single triplet of pairs having the same $f_{homo} = 0.4$, 0.6, 0.65, or 0.75; one f_{homo} for one triplet for each plot. For the noisiest triplet when $f_{homo} = 0.4$ (top plot of Fig. 4B), the average accuracy was 0.77 and 0.71, respectively, for the more and less congruent RDSs in a pair, and the signal preference P for the more congruent pair was insignificant in each RDS pair, and remained so even when the 3×52 trials for all the three RDS pairs of this triplet were pooled to get $\bar{P}$ for stronger statistics. For the other three triplets, $f_{homo} \geq 0.6$, the average accuracy was no less than 0.91, and the individual accuracy (for each RDS) was no less than 0.85 (this worst case was for the neutral RDS in the congruent-neutral pair in the second plot in Fig. 4B), and the average preference $\bar{P}$ for the more congruent RDS was significant when all the 3×52 trials were pooled for each triplet. For the two least noisy triplets (the bottom two plots of Fig. 4B) when $f_{homo} = 0.65$ and $f_{homo} = 0.75$, respectively, the preference $P_{\text{C-I}}$, contributed by both the pro- and anti-effects of the reversed depth signals, was significant even though only 52 trials (of the congruent-incongruent pair) were involved.

It is not surprising that the preference for the more congruent RDS is most easily manifested when the RDSs are not too noisy and so yield sufficiently clear depth perception. Similar dependence on the accuracy of the depth reports of the reversed signal effect was also seen in other observers. Since different observers required different f_{homo} values to see the depth order clearly, we adjusted f_{homo} values individually for individual observers so that each observer could achieve an accuracy (of depth reports) of no less than 80% for each RDS in any non-control RDS pair within a single session in Experiment 1. In the rest of the results reported for Experiment 1, only data from sessions satisfying this requirement are included.

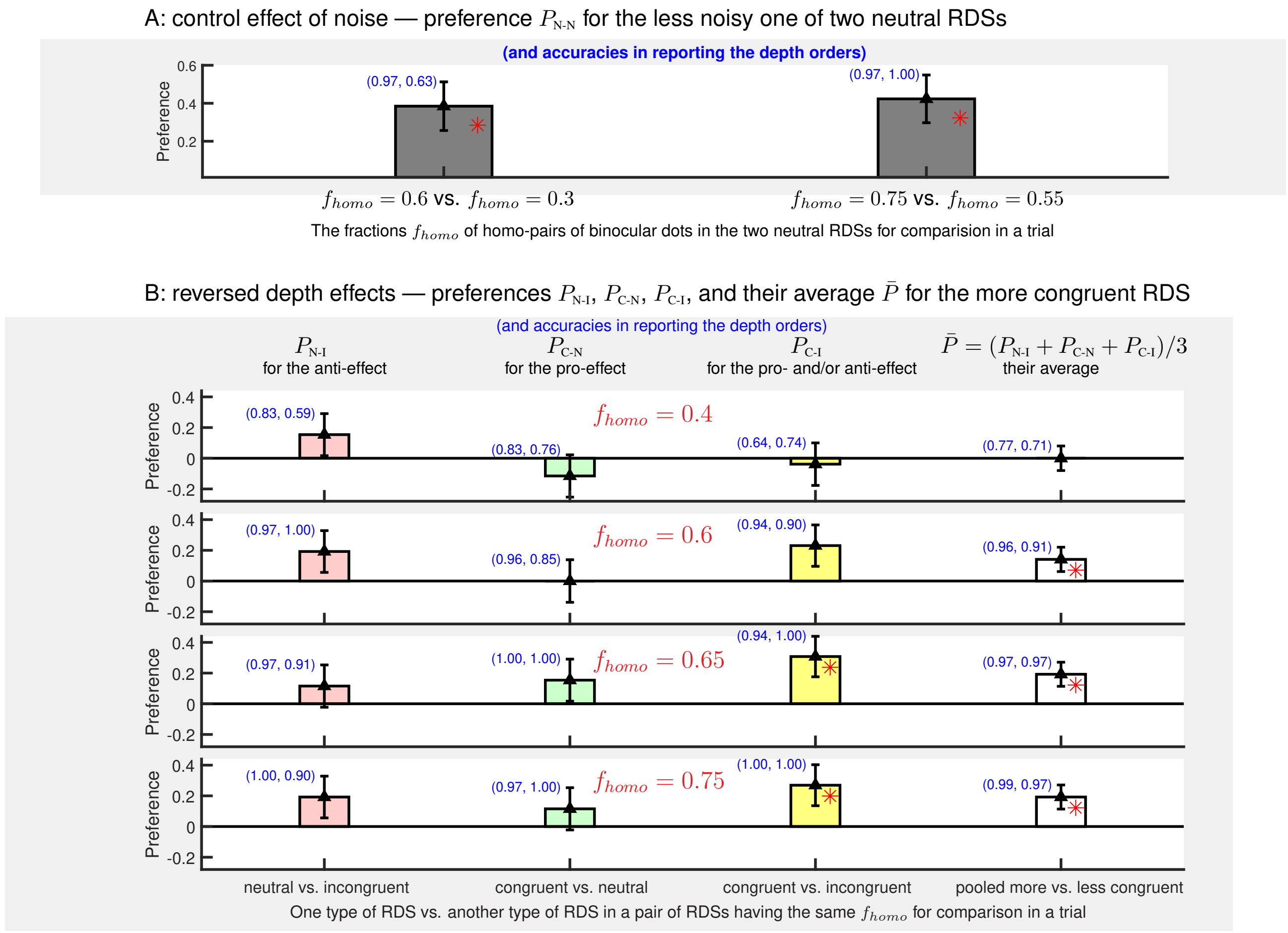

Figure 4: The perceptual effect of the reversed depth signals is weaker or absent when the RDS is too noisy for accurate depth discrimination, shown by an example naive observer in two experimental sessions with viewing duration T = 0.02 second in Experiment 1. In each trial, the observer viewed two RDSs and reported which RDS appeared clearer in the perceived depth order between the central disk and the surrounding ring, and then reported the depth order in the reported RDS. A: when both RDSs are neutral (their f_{homo}'s indicated on the horizontal axis), preference $P_{\text{N-N}}$ (defined in equations (2-6)) for the less noisy RDS (which had a larger f_{homo}) is significant (indicated by red '*'). The blue-colored numbers next to each data point are accuracies for depth order reports (the left number for the less noisy, or more congruent in B, RDS). B: when the two RDSs in a trial differed in congruency, preferences $P_{\text{N-I}}$ (pink bar, for the anti-effect), $P_{\text{C-N}}$ (green bar, for the pro-effect), $P_{\text{C-I}}$ (yellow bar, for the pro- and/or anti-effects), or their average $\bar{P}$ for the more congruent RDS was significant only when the f_{homo} for the RDSs was large enough to allow accurate reports of the depth orders. Each plot, formatted similarly to A, is for a triplet of RDS pairs with a given $f_{homo} = 0.4$, 0.6, 0.65, or 0.75 as indicated. In each plot, the left three (pink, green, and yellow) data bars are $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$, respectively, for individual RDS pairs (marked on the horizontal axis) of the triplet, the white bar on the right is $\bar{P}$, the average of the left three bars. All the error bars are the estimated standard errors calculated as described in section (2.4). Each non-neutral RDS had $f_{hetero} = 0.2$. The first two plots in B and the left data bar in A are from trials of one experimental session, the bottom two plots in B and the right data bar in A are from the other session.

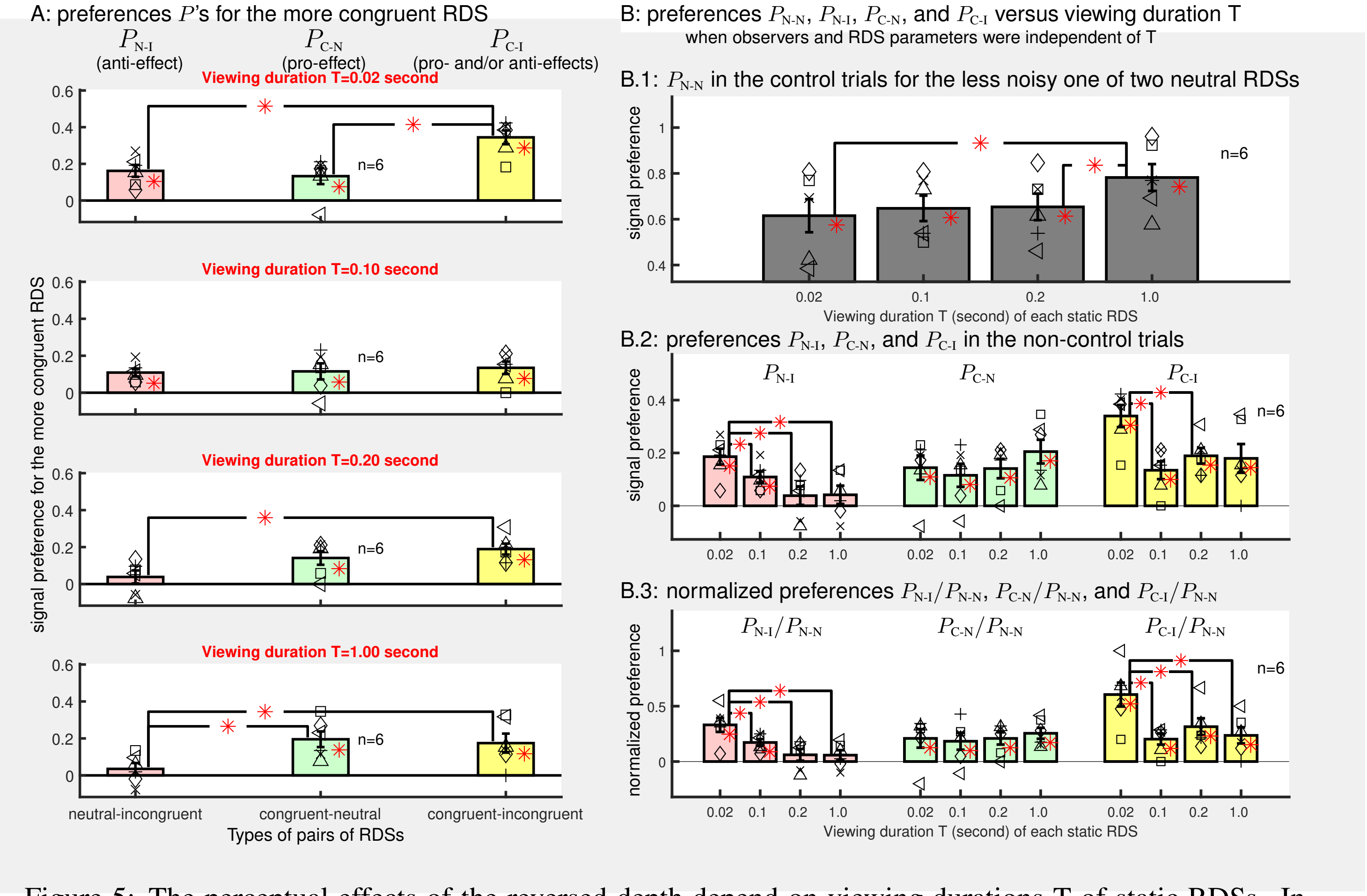

Figure 5: The perceptual effects of the reversed depth depend on viewing durations T of static RDSs. In particular, the anti-effect $P_{\text{N-I}}$ became insignificant for longer T. Data from sessions (of Experiment 1) in which accuracies of depth order reports for non-control RDS pairs were no less than 80%. Data from different observers are marked by differently shaped symbols $+$, $\times$, $\diamond$, $\triangle$, $\square$, $\triangleleft$. Symbol $\square$ is for the only non-naive observer, the author; symbol $\triangle$ is for the observer in Figure 4. A: preferences $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$ for the more congruent RDS in a pair. The RDSs in each pair (indicated on the horizontal axis) had the same fraction f_{homo} of homo-pairs of dots. Each plot is for one viewing duration T= 0.02, 0.1, 0.2, or 1.0 second. B: preferences versus the viewing duration T using only data from experimental sessions in which the set of RDS stimulus parameters (f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), f_{homo}(triplet 2)) depended only on the observer but not on viewing duration T. B.1: preference $P_{\text{N-N}}$ for the less noisy RDS in the control trials involving two neutral RDSs. B.2: preferences $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$ for the more congruent RDS in the non-control trials in the same data session as that in B.1 for each observer. B.3: normalized preferences $P_{\text{N-I}}/P_{\text{N-N}}$, $P_{\text{C-N}}/P_{\text{N-N}}$, and $P_{\text{C-I}}/P_{\text{N-N}}$. In A and B, all the non-neutral RDSs had $f_{hetero} = 0.2$. All stimulus dots were $0.174^o \times 0.174^o$ squares. Each bar is an average of the corresponding quantities across observers, the error bar is the standard error of this average. A red '*' in a bar indicates that this observer-averaged quantity is significantly larger than zero, a red '*' between two bars linked by black lines indicate that the two averages are significantly different from each other by a paired t-test (this significance is not assessed in B between two differently-colored data bars).

3.1.2 The reversed depth signals impact depth perception along the direction of V1 responses

Using data from experimental sessions that satisfied the accuracy requirement (that the accuracy for depth reports was no less than 80% for every RDS in all the non-control RDS pairs), the top plot of Fig. 5A shows the signal preferences $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$ for the more congruent RDS in six observers, and the averages across these observers, for a viewing duration T = 0.02 seconds. These results include the data from the less noisy session for the example observer in Fig. 4. Each preference $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$ for each observer (visualized by the observer-specific symbol) comes from pooling 2×52 trials from two RDS pairs (except for the observers noted in this footnote [1]) of the type neutral-incongruent, congruent-neutral, or congruent-incongruent, respectively, from two triplets that differed in f_{homo} in the same session (see Methods). For example, the left-most data point $\triangle$ comes from pooling the trials from the left-most data bars in the third and fourth plots of Fig. 4. Averaged across the observers, each preference $P_{\text{N-I}}$, $P_{\text{C-N}}$, or $P_{\text{C-I}}$ is significant. Furthermore, the preference $P_{\text{C-I}}$, contributed by both the pro- and anti-effects, is significantly larger than both the preference $P_{\text{N-I}}$ for the anti-effect and the preference $P_{\text{C-N}}$ for the pro-effect. These data demonstrate significant impact of the reversed depth signals on perceived depth for T = 0.02 second. This T is likely sufficiently brief that the top-down feedback processes are not fully effective (see Discussion).

The other plots of Fig. 5A show analogous results for longer viewing durations T = 0.1, 0.2, and 1 second. In each case, at least two of the three observer-averaged preferences, $P_{\text{N-I}}$, $P_{\text{C-N}}$, and $P_{\text{C-I}}$, are significant.

3.1.3 The anti-effect of the reversed depth signals is absent for longer viewing durations

To examine the effect of viewing duration T more closely, the individualized set of RDS parameters, f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), f_{homo}(triplet 2), was fixed for each observer across multiple sessions for different durations T = 0.02, 0.1, 0.2, and 1.0 second. Fig. 5B.1 indicates that observers' preferences $P_{\text{N-N}}$ for the less noisy RDS in the control pair of two neutral RDSs increased somewhat with T. This is not surprising since a longer viewing duration gives observers more time to integrate sensory signals for discrimination. However, the preference $P_{\text{N-I}}$ for the anti-effect decreased with T. Meanwhile the preference $P_{\text{C-N}}$ for the pro-effect did not vary significantly with T. Unsurprisingly, the preference $P_{\text{C-I}}$ manifesting both the pro- and anti-effects also decreased with T (Fig. 5B.2). Fig. 5B.3 replots the results of Fig. 5B.2 by normalizing the preferences for the more congruent RDSs according to the preference for the less noisy neutral RDS of the same observer in the same data session (same T), and reveals qualitatively the same outcome. These results suggest that the pro-effect of the reversed depth signals is not vetoed by the feedback verification. In contrast, the anti-effect is vetoed by the feedback when the viewing duration is sufficiently long to make the feedback fully effective.

[1] In two instances, a given duration T yielded two data sessions for a single observer satisfying the accuracy requirement for depth reports. In each instance, the two sessions differed in the set of RDS parameters (f_{homo}(control), Δf_{homo}(control), f_{homo}(triplet 1), f_{homo}(triplet 2), see Methods), and the reported preferences for this observer and this duration T are then the averages across the two sessions. One instance involved the non-naive observer (with data symbol □) for T=0.02 second and the other involved a naive observer (data symbol ◁) for T = 1 second. The qualitative results in Fig. 5A remain unchanged whether or not these two additional sessions of data are included in the data analysis and regardless of which one of the two sessions for a given observer and given T is included in the data analysis.

3.1.4 Reducing the dot size increases the anti-effect of the reversed depth signals

One might imagine that smaller stimulus dots could make it more difficult for top-down feedback to verify whether the binocularly corresponding dots in the actual inputs are matched rather than mismatched. If so, with smaller dots the feedback should be less likely to veto the reversed depth signals from the hetero-pairs of dots. Fig. 6, which is arranged as in Fig. 5, shows that this is indeed the case. When the dot size (its area) was reduced by a factor of four, by halving the side length of the square-shaped dot (while increasing the dot density by a factor of four), the pro-effect stayed unchanged whereas the anti-effect became stronger. Naturally, the $P_{\text{C-I}}$ manifesting both the pro- and the anti-effects also became stronger. Meanwhile, this dot size reduction did not cause any significant change in the control preference $P_{\text{N-N}}$ for the less noisy RDS among two neutral RDSs in control trials, see Fig. 6B.1.

Qualitatively, a reduction in dot size has a very similar impact on the preferences to a reduction in the viewing duration. This is consistent with the idea that both manipulations made feedback verification less effective.

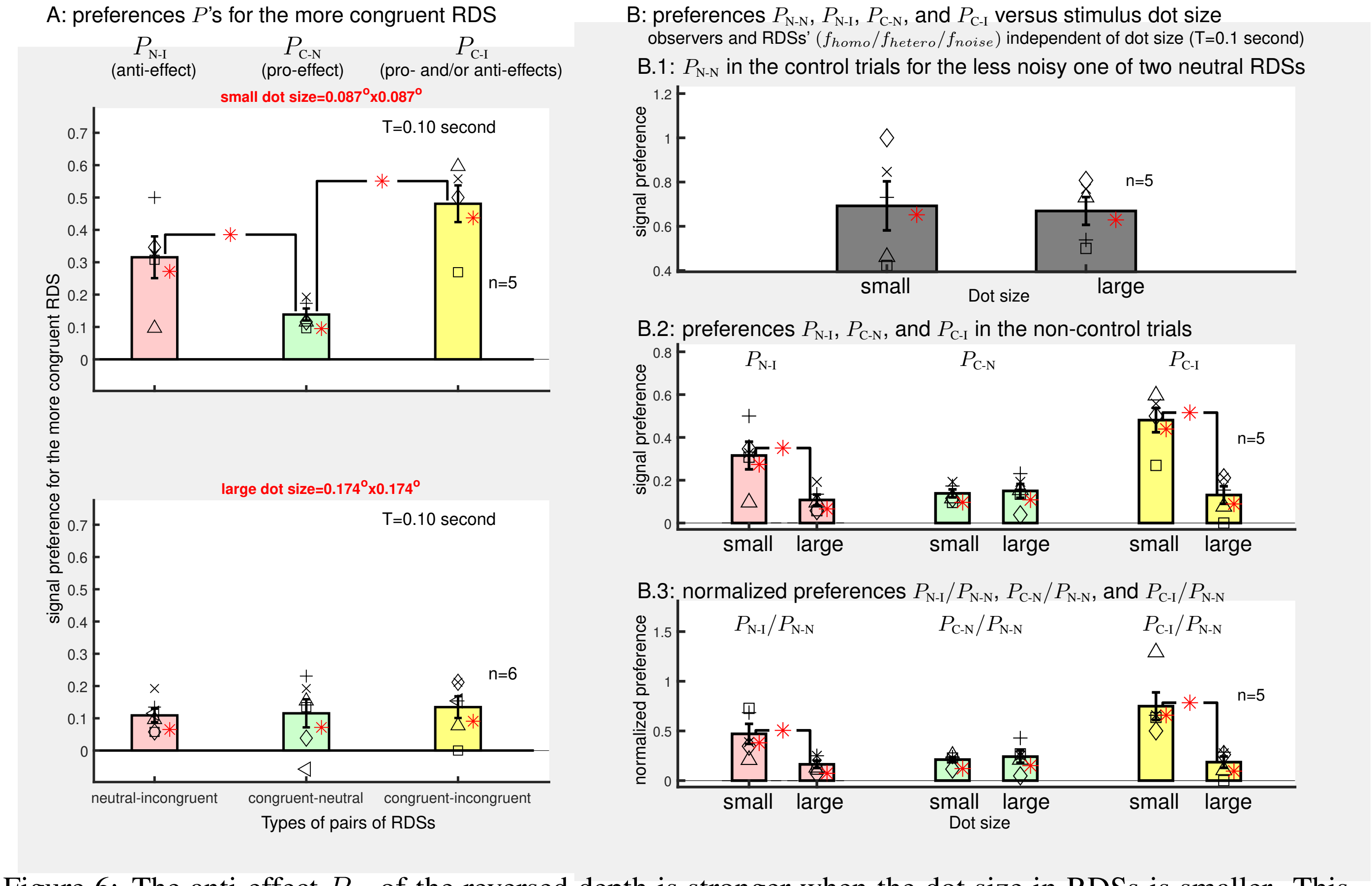

Figure 6: The anti-effect $P_{\text{N-I}}$ of the reversed depth is stronger when the dot size in RDSs is smaller. This figure uses the same requirements on the accuracies for the depth order reports for data inclusion, the same observer visualization by data symbols, and the same plotting format as that in Fig. 5. All the error bars represent the standard errors of the observer averages. The stimulus dot density (dots/per unit image area) was scaled inversely to the area size of each dot to keep the image area covered by the dots statistically unchanged by dot size changes. All the non-neutral RDSs had $f_{hetero} = 0.2$.

3.2 Experiment 2: threshold f_{homo} amount of homo-pairs of dots needed to discriminate depth orders in RDSs

Experiment 2 assessed the effect of reversed depth signals by quantifying their impact on the minimum fraction f_{homo} of homo-pairs of dots needed in order to discriminate the depth order between the disk and the ring. We call this minimum fraction the threshold. Experiment 1 showed that reversed depth signals could increase or decrease the clarity of perceived depth when they were congruent or incongruent, respectively. If these superthreshold effects generalize to threshold level behavior in Experiment 2, then one expects that the reversed depth signals should decrease or increase the threshold f_{homo} when they are congruent or incongruent, respectively.

However, Fig. 4 suggests that the perceptual difference due to reversed depth signals was weak or absent when observers could not discriminate the depth order accurately. This was particularly so for T = 0.02 second. Hence, if the threshold f_{homo} was not affected by reversed depth signals when T = 0.02 second, we could not be sure whether this would be because observers did not have sufficient viewing time to integrate the depth signals sufficiently. Experiment 2 improved over Experiment 1 by including an additional stimulus class: dynamic RDSs made of consecutive dichoptic image frames, each for 0.02 second, for a total duration of 0.2 second. Different dichoptic image frames within the 0.2 second depicted the same 3D scene (the disk and the ring with their respective disparities) using the same fractions f_{homo} and f_{hetero} of the homo- and hetero- pairs of dots for the RDS in the trial (see Methods). Meanwhile, the random set of stimulus dots in each frame was generated independent of the random set in any other frames (and frames in any other trials). If 0.02 second is too short a time to allow the feedback verification of the stimulus dots to occur before those dots were replaced by another random set of dots in the next dichoptic frame, then the veto of the reversed depth signals in the dynamic RDS would be less effective than that in a static RDS viewed for the same overall duration of T = 0.2 second.

Threshold values of f_{homo} were obtained by the staircase procedure detailed in the Method section. During an experimental session, each trial showed a RDS which was randomly one of six (3×2) types of RDSs ((neutral, congruent, or incongruent) $\times$ (static or dynamic)). The central disk was randomly in front or behind the surrounding ring in each trial. In each trial, the observer had to provide a forced choice report whether the central disk was in front or behind the surrounding ring. The f_{homo} values for the six types of RDSs were adjusted in parallel and independently by a staircase method. Meanwhile, $f_{hetero} = 0.3$ was fixed for all the non-neutral RDSs (except in very rare trials which were detailed in the Method section and not included for the threshold calculation). Figure 7 shows that indeed the thresholds f_{homo} were significantly lower when the RDS was congruent than when it was neutral. This is so for both the dynamic and static RDSs. However, the threshold f_{homo} was significantly larger in incongruent than in neutral RDSs only for the dynamic RDSs. Overall, the difference between the threshold f_{homo} for the incongruent RDSs and that for the congruent RDSs was significantly larger for the dynamic than the static RDSs. These results are in line with those in Fig. 5. In particular, the pro-effect is not affected by whether the RDS is dynamic or static, whereas the anti-effect is insignificant for the static RDS but significant for the dynamic RDS.

When the RDS is dynamic and congruent, Figure 7A shows that the threshold f_{homo} was close to

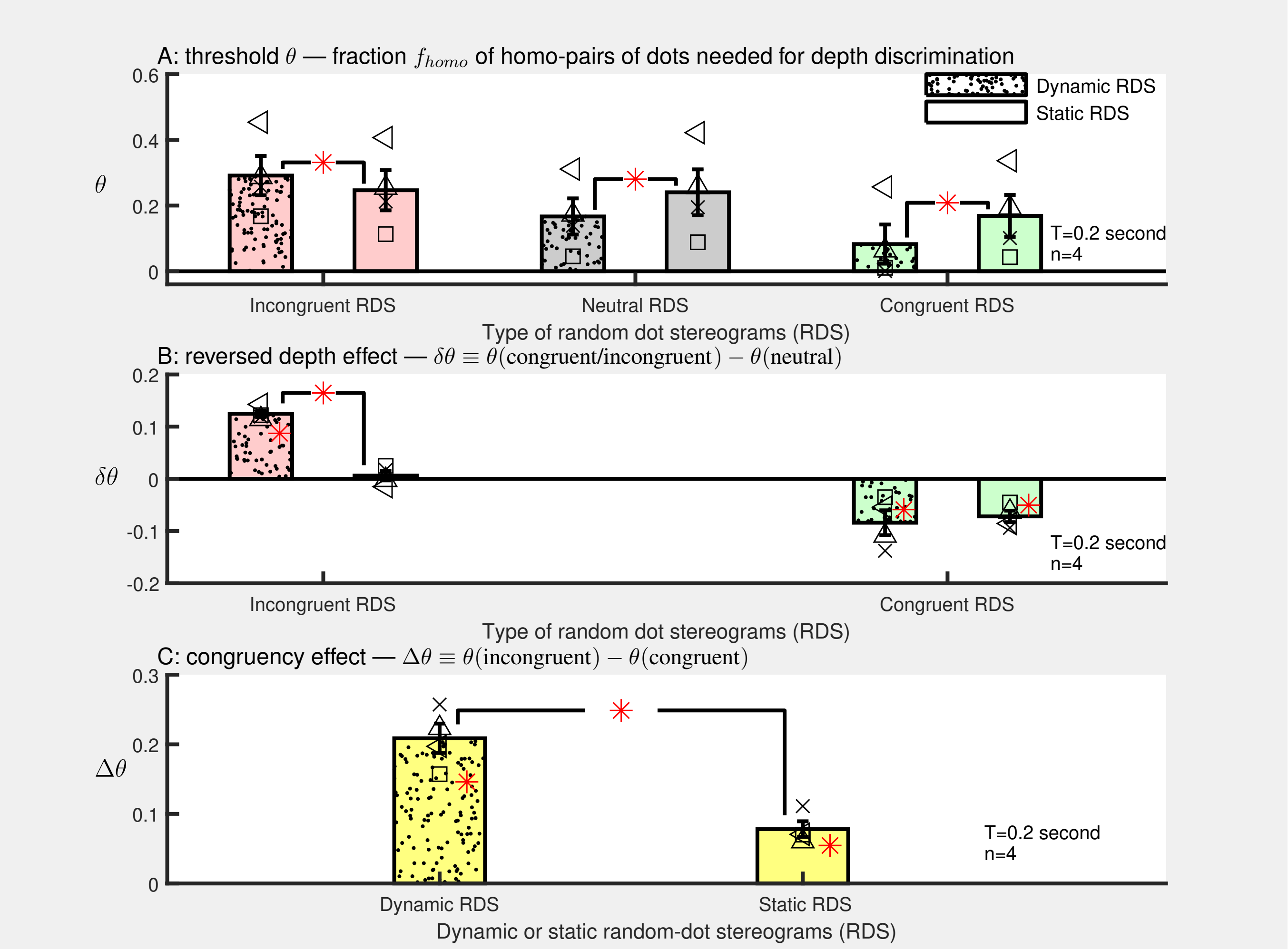

Figure 7: Threshold fraction f_{homo} of homo-pairs of dots needed to see the depth order of a RDS (viewed for 0.2 second in each trial of Experiment 2) varied with the congruency (incongruent, neutral, or congruent) of the RDSs, more so for the dynamic RDSs. A: this threshold, denoted as θ, in $n = 4$ observers obtained by a staircase method. In each trial, the RDS was randomly either static or dynamic (in which the set of random dots was replaced every 0.02 second by another, independently generated, set while keeping the other stimulus characters, including f_{homo}, f_{hetero}, and f_{noise}, unchanged), and was randomly incongruent, neutral, or congruent. B: the difference $\delta\theta$ between the threshold for non-neutral RDSs and that for neutral RDSs in A. C: the difference $\Delta\theta$ between the threshold for incongruent RDSs and that for congruent RDSs in A. Definitions of the data symbols for individual observers are as in the previous figures. Each bar is the average of the corresponding quantities across the observers, the error bar is the standard error of this average. In B and C, a red '*' on a data bar indicates that this observer-averaged quantity is significantly different from zero. A red '*' between two same-colored bars linked by black lines indicates a significant difference (by paired t-test) between the two averages. All non-neutral RDSs had $f_{hetero} = 0.3$.

zero for two observers, implying that these two observers were seeing depth orders sufficiently well relying mostly or only on the reversed depth signals (i.e., with only very little normal depth signal). Hence, reversed depth could be visible in central vision, albeit perhaps at threshold level performance, using dynamics RDSs in which each frame duration is 0.02 second or shorter. This can be investigated further (Zhaoping, in preparation).

Interestingly, for neutral RDSs without any reversed depth signals, the threshold f_{homo} is lower in dynamic than static RDSs, see Figure 7A. Perhaps the degradation of the feedback verification also helped to mitigate the effect of noise.

4 Summary and Discussion

4.1 Summary of experimental findings

Contrary to the case in the peripheral visual field(Zhaoping and Ackermann, 2018), reversed depth signals from hetero-pairs of dots in random dot stereograms (RDSs) are typically invisible in the central visual field. Our study shows that, when these invisible reversed depth signals are mixed with normal depth signals from homo-pairs of dots in RDSs, they can impact the quality or strength of the perceived depth. Specifically, in a RDS made noisy by the presence of binocularly non-corresponding noise stimulus dots, the depth order associated with normal depth signals can be more clearly perceived when the reversed and normal depth signals agree on the depth order. This enhanced perceptual clarity was demonstrated (in Experiment 1) by comparing two noisy RDSs that have the same amount of normal depth signals and differ in the presence of the congruent reversed depth signals. Additionally (Experiment 2), the presence of the congruent reversed depth signals reduces the threshold amount of normal depth signals (i.e., the amount of homo-pairs of dots) needed in order to discriminate the depth order. When the reversed depth signals are incongruent such that they disagree with the normal depth signals on the depth order, they make the normal depth order less clearly perceived, but only when each RDS image was sufficiently brief. This is demonstrated in both Experiments 1 and 2 in manners analogous to that for the congruent signals. The pro-effect associated with congruent signals does not depend sensitively on the viewing duration of the RDS or the size of the stimulus dots. The anti-effect from the incongruent signals is reduced as the RDS viewing duration increases from 0.02 to 0.1 seconds and is eliminated by longer viewing durations; and it is enhanced when the RDSs have smaller dot sizes.

4.2 Support of the Feedforward-feedback-verify-reweight (FFVW) process for central vision

The experimental findings support the feedforward-feedback-verify-reweight (FFVW) process for visual inference(Zhaoping, 2019). Due to the attentional bottleneck which starts from V1's output, only a limited amount of visual input information is sent forward to higher visual areas(Zhaoping, 2019). In particular, information about the eye-of-origin of visual inputs and some spatial details of visual inputs are not sent forward. In ambiguous and noisy conditions, top-down feedback can aid visual recognition by querying for more information via the FFVW process. Such feedback verifies the perceptual hypotheses suggested by the feedforward signals from V1 by sending back the synthesized would-be inputs for each hypothesis and checking whether they match the actual sensory inputs. In ambiguous or noisy situations, typically multiple hypotheses are suggested by the feedforward signals. For example, a noisy RDS could send forward three possible hypotheses: (1) disk in front; (2) disk behind; and (3) no depth difference between the disk and the ring, with three different weights according to the strengths and characteristics of V1's feedforward signals. The feedback verification and reweighting in FFVW modifies these three weights according to the match between the would-be inputs for each hypothesis and the actual inputs. When hetero-pairs and noise are the actual inputs (without homo-pairs), one feedforward hypothesis is the depth order, e.g., front, according to the reversed depth signals. Its would-be inputs are homo-pairs with a disparity opposite to that in the hetero-pairs of dots. These would-be inputs are synthesized according to our brain's internal model of the visual world. This internal model has likely learned from visual experience that an object in the world typically forms contrast-matched (e.g., homo-pairs), rather than contrast-reversed (e.g., hetero-pairs), images in the two eyes. Hence, the suggested depth order is vetoed by the feedback verification since the actual hetero-pairs and the would-be homo-pairs do not match. This veto diminishes the weight for this depth order, making it invisible to perception in the central vision. Peripheral vision can perceive this depth order by the reversed depth signals(Zhaoping and Ackermann, 2018) as predicted by CPD, due to a lack of feedback to veto the feedforward perceptual hypothesis.

When the sensory inputs additionally contain homo-pairs whose normal depth signals suggest the same depth order, e.g., front, as the reversed signals by the hetero-pairs, this is a congruent RDS. The would-be inputs in the feedback can find their matches in the homo-pairs in the sensory inputs. This match helps to confirm the initial feedforward depth order to make it perceptible. Equivalently, one may view the situation as follows. The homo-pairs evoke the feedforward hypothesis, e.g., front; the hetero-pairs evoke the same hypothesis, and the would-be inputs in the feedback find their matches in the homo-pairs to make this depth order perceptible. Would this make the hetero-pairs irrelevant as visual inputs because the reversed signals are redundant? The answer is negative according to our data. The perception of this depth order is clearer with the congruent hetero-pairs than the perception when the hetero-pairs of dots are replaced by noise dots. This implies that the reversed depth signals from the hetero-pairs are not treated as noise, but instead added to the normal depth signals from the homo-pairs to increase the feedforward weight for the common depth order. This enhanced feedforward weight leads to an enhanced weight at the eventual perceptual outcome, yielding the pro-effect (Fig. 3).

When the normal depth signals suggest a depth order, e.g., front, that is opposite to the depth order (e.g., back) suggested by the reversed depth signals, the RDS is incongruent. The feedforward signals from V1 to higher brain areas contain at least two conflicting hypotheses about the depth order: one is front; the other is back. The would-be inputs for the former can match well with the actual inputs of the homo-pairs, but the would-be inputs for the latter can match with neither the homo-pairs nor the hetero-pairs. Hence, the depth order (back) suggested by the reversed depth signals would be vetoed and made invisible perceptually. Our data for a static RDS viewed for longer than 0.1 second suggest that such a veto made the hetero-pairs of dots perceptually treated as if they were merely noise dots. For example, a disk in an incongruent RDS with 30% of its dots from homo-pairs, 30% from hetero-pairs and the rest from noise dots appeared in our data equivalent to a neutral RDS with 30% of its dots from homo-pairs and 70% from noise dots. However, when the viewing time was shorter (e.g., 0.02 second for a static RDS or for a single frame in a dynamic RDS), this incongruent RDS appeared in our data as being noisier, or less clear, than the neutral RDS in depth order perception. This is consistent with the idea that when there is insufficient time for the feedback in the FFVW process, the conflicting hypothesis of the depth order by the hetero-pairs is not effectively vetoed, enabling it to compete with, and thus weaken, the normal depth order by the homo-pairs for perceptual outcome. This competition and weakening give rise to the anti-effect by the reversed depth signals (Fig. 3).

To veto the reversed depth order, the feedback verification in the FFVW process must have a sufficient spatial resolution to identify the contrast reversal between binocularly corresponding dots in the hetero-pairs. If so, smaller dots should make this verification more difficult, and thus the anti-effect of the reversed depth in an incongruent RDS should be stronger in RDSs with smaller dots. This is consistent with our data in Fig. 6. Meanwhile, this weakened verification does not affect the pro-effect by the reversed signals that are congruent. This is consistent with our data in Fig. 5 suggesting that the pro-effect is not significantly affected by more or less feedback enabled by a longer or shorter viewing duration.

Manipulating the viewing duration and dot size affected the anti-effect of the reversed depth signals while leaving the pro-effect more or less unchanged (see Fig. 5A, Fig. 6A, and Fig. 7B). This implies that the feedback verification in the FFVW process, when made effective, mainly or exclusively corrected the incongruent rather than the congruent part of the reversed depth signals fed forward from V1. This is a nonlinearity in the perceptual inference process: the erroneous reversed depth signals are utilized constructively when they are congruent with the perceptual outcome according to some other sensory inputs but ignored when they are incongruent. Such a nonlinearity is characteristic of analysis-by-synthesis computation, often seen in phenomena such as sensory filling-in and (modal or amodal) completion by input contexts(Zhaoping, 2014; Zhaoping and Jingling, 2008). This filling-in effect helps the binocular mismatch in the hetero-pairs to be disregarded or downplayed, while allowing the reversed depth signals generated by these pairs to boost the overall signal.

4.3 Relation with previous works

4.3.1 Contrast-reversed simple stereograms or complex RDSs in the central and peripheral visual field

In the central visual field, humans can see disparity-defined veridical depth in simple contrast-reversed stereograms (when an item is bright in one eye and dark in the other eye) containing only one or a few items(Helmholtz, 1925; Cogan et al., 1995). They can also see the veridical depth in contrast-reversed RDSs when the dot density is much lower than those used in physiological experiments to evoke V1 responses(Cumming and Parker, 1997; Cumming et al., 1998) or those in the current study. These results suggest that human also have other cues, perhaps from vergence, that allow them to see the veridical depth in contrast-reversed stereograms, and presumably such cues are ineffective for dense contrast-reversed RDSs. Humans exhibit an apparent and weak perception of the reversed depth by small disparities in central vision when visual inputs are restricted to vertical gratings within a narrow spatial frequency band(Read and Eagle, 2000). It is likely that this perception is by using normal rather than reversed depth signals, since a dichoptic pair of one grating and its contrast-negative version by a disparity d less than half of the grating's wavelength λ is equivalent to another dichoptic pair of one grating and its contrast-matched version by a disparity $d' = \lambda/2 - d$ in the opposite direction. Indeed, the perception of this reversed depth was reduced when the bandwidth of the grating was increased(Read and Eagle, 2000).

However, it has been known for many years that, in dense and contrast-reversed RDSs placed in the central visual field, humans can perceive neither the veridical depth nor the reversed depth reported by V1 neurons(Cumming et al., 1998; Hibbard et al., 2014; Asher and Hibbard, 2018). Most of these observations were made using RDSs viewed for at least 0.1 second for each image frame. Neri et al. (1999) showed an indirect perceptual effect of the reversed depth signals by an aftereffect on the normal depth perception after adapting on contrast-reversed RDSs, and this aftereffect is in the direction as if the observers had adapted to normal depth signals congruent with the reversed depth signals.

In some previous reports(Doi et al., 2011, 2013; Aoki et al., 2017; Tanabe et al., 2008) of weak or moderate degrees of reversed depth perception, the depth surface was in fact centered around $3^o - 5.5^o$ eccentricity rather than the fovea. Based on CPD, it is likely that at such eccentricities the top-down feedback verification is less effective, making the reversed signals more likely to be visible. Additionally, in various of these studies(Tanabe et al., 2008; Doi et al., 2013; Aoki et al., 2017), some stimuli were dynamic RDSs with each RDS image frame renewed every 0.023 second. As discussed in this paper, such a short duration is also likely to make the feedback in the FFVW process less effective. Zhaoping and Ackermann (2018) showed robust reversed depth perception using a disk centered at 10.1^o eccentricity surrounded by a ring. Their RDSs were dynamic with a duration of 0.1 second (similar to that in Doi et al. (2011), but the RDSs were displayed using a mirror stereoscope rather than shutter-goggles). The reversed depth perception was robustly present at this eccentricity across various stimulus variations in the sizes of the dot, the disk, the ring, and the disparity step and in whether a gap was present between the disk and the ring. However, when the disk was centered around 1.8^o-4.1^o for the same observers, the reversed depth was not perceived (or too weak to reach significance)(Zhaoping and Ackermann, 2018). Inconsistencies between different reports on the presence or absence of the perception of the reversed depth in parafoveal vision has been noted previously(Hibbard et al., 2014)

However, none of these previous studies investigated reversed depth signals that are congruent with simultaneously presented normal depth signals.

4.3.2 Neural mechanisms behind the generation and processing of reversed depth signals

Opposite disparity tuning curves to contrast-reversed dichoptic inputs can be understood in terms of the disparity energy model of V1 complex cells(Ohzawa et al., 1990; Qian, 1994; Cumming and Parker, 1997). A classic complex cell is made of a quadrature pair of two simple cells, each simple cell $i = 1$ or 2 applies a dichoptic pair of filters, $f_{i,l}(x)$ and $f_{i,r}(x)$, for the left and right eye's receptive fields, respectively, as a function of space x, to a dichoptic pair of visual inputs, $I_l(x)$ and $I_r(x)$, for the two eyes. Let $L_i \equiv \int f_{i,l}(x) I_l(x) dx$ and $R_i \equiv \int f_{i,r}(x) I_r(x) dx$, the complex cell's output response is

$$O \equiv (L_1 + R_1)^2 + (L_2 + R_2)^2 = L_1^2 + L_2^2 + R_1^2 + R_2^2 + 2L_1 \cdot R_1 + 2L_2 \cdot R_2 \quad (9)$$

$$\approx \text{constant } + 2L_1 \cdot R_1 + 2L_2 \cdot R_2. \quad (10)$$

The constant $\approx L_1^2 + L_2^2 + R_1^2 + R_2^2$ arises from a quadrature structure between $f_{1,l}/f_{1,r}$ and $f_{2,l}/f_{2,r}$ (Qian, 1994; Qian and Mikaelian, 2000), and this constant is invariant to the disparity between inputs $I_l(x)$ and $I_r(x)$. Consequently, the disparity tuning largely arises from $2L_1 \cdot R_1 + 2L_2 \cdot R_2$, which is effectively the correlation between the two monocular inputs through the lens of the receptive field filters. This correlation promptly inverts sign when one of the monocular image, e.g., $I_r(x)$, inverts its contrast to make $R_i \rightarrow -R_i$, so that, in the disparity tuning curve of the outcome O versus disparity, a peak or trough (for preferred or non-preferred disparity) becomes instead a trough or peak, respectively.

However, unlike cells excited by a contrast-matched stereogram, complex cells excited by a contrast-reversed depth surface do not share a common preferred disparity(Read and Eagle, 2000; Asher and Hibbard, 2018). For example, if the contrast-reversed stereogram contains a near disparity, then the excited neurons prefer different far disparities, with some preferred disparities being further than others. This is because different V1 neurons prefer different spatial frequencies and the preferred disparities are to some degree scaled inversely with the preferred frequencies(Zhaoping, 2014). Hence, the activated neurons largely agree on the qualitative depth order, near or far, but not on the quantitative depth magnitude. Such V1 signals are sufficient for the task in the current and previous psychophysics studies, if they are well utilized by subsequent brain processes, although they are likely to make it difficult to perceive a coherent depth surface.

To contrast-reversed RDSs, neurons in higher visual area V4 of monkeys are much less tuned to disparity compared to V1 neurons(Tanabe et al., 2004). A similar hierarchy is observed in visual Wulst of owls, where neurons less tuned to disparity in contrast-reversed RDSs tend to have longer response latencies(Nieder and Wagner, 2001) (which suggests that these neurons might be at a location further downstream along the visual processing pathway). This hierarchical progression of the representation of depth signals does not indicate clearly whether the underlying processes involve any feedback. Meanwhile, some models(Lippert and Wagner, 2001) suggest that feedforward processes are sufficient for such a hierarchical representation of depth signals.

4.3.3 Feedforward models of processing of the random dot stereograms

Various feedforward and phenomenological models have been suggested for depth or disparity processing. Specifically, these models use some linear and nonlinear transforms, such as filtering, squaring, rectification, and binocular matching, of the sensory input signals without involving feedback or recurrent processes. The outcomes of these feedforward transforms are used to model the neural or behavioral responses.

Henriksen et al. (2016) modified the classical energy model so that the responses of V1 complex cells is the square of the outcome O from the energy model in equation (9). This squaring of the O adds a nonlinear transform to the binocular correlation such that positive correlation (e.g., in homo-pairs) is weighted more than negative correlation (e.g., in hetero-pairs). They used this model to explain why V1 neurons' disparity tuning is weaker to contrast-reversed than contrast-matched RDSs, and why V1 cells remain weakly tuned to disparity in half-matched RDSs (that have 50% of the dots from homo-pairs and 50% from (incongruent) hetero-pairs) without reversing the disparity preference. They used this finding to explain human perception of normal depth orders in (incongruent) half-matched RDSs(Doi et al., 2011). We predict that disparity tuning in the V1 cells, and likely also in cells in higher visual areas, to half-matched RDSs should be stronger when responding to congruent rather than incongruent RDSs.

A stronger nonlinear operation than the squaring of O by Henriksen et al. (2016) is to apply a threshold rectification on the output O (in equation (9)) of the energy model(Lippert and Wagner, 2001; Nieder and Wagner, 2001). Equation (9) indicates that this threshold on the energy model outcome O is like a threshold on the binocular correlation (through the lens of the dichoptic receptive fields). When the threshold is such that the positive correlation from the homo-pairs is above the threshold whereas the negative correlation from the hetero-pairs is below the threshold, this threshold energy model is then equivalent to the cross-matching model by Doi and Fujita (2014). When each dichoptic receptive field is small enough to contain no more than one stimulus dot in RDSs, the cross-matching model requires binocular matching of individual stimulus dots to activate depth signal detectors.

Instead of the *binocular correlation* only as suggested by the energy model, or the *binocular matching* only by the extreme cross-matching model, Doi et al. (2011) and Fujita and Doi (2016) proposed that depth perception arises from a weighted summation of these two mechanisms. To explain their behavioral data(Doi et al., 2011, 2013), the two weights for the two mechanisms are required to be adjusted to suit different input conditions. In particular, the weights for the correlation and matching mechanisms are increased and decreased, respectively, for inputs that are more transient (or of higher temporal frequency) or having a larger magnitude of disparity steps.

Asher and Hibbard (2018) analyzed the first- and second-order mechanisms for depth processing. The first order mechanism is defined as in the disparity energy model in equation (9). The second order mechanism is the same as the first order mechanism except that the two monocular input images, I_l and I_r, are images of spatial (band-pass filtered) luminance contrast rather than the original input luminance images. The structure of the second-order mechanism is motivated by observations showing that normal depth can be perceived behaviorally and detected by neurons (with disparity tuning) in cat area 18 when the underlying binocular matching was based on envelops of luminance contrast rather than luminance(Wilcox and Hess, 1996; Tanaka and Ohzawa, 2006). The normal (non-reversed) depth perception in simple contrast-reversed stereograms with a few object items in the image(Cogan et al., 1995) can be viewed as a special case of this. Contrast-matched and contrast-reversed stereograms appear the same to the second-order mechanism, which is thus very different from binocular matching. Asher and Hibbard (2018) explained that humans cannot perceive depth in a contrast-reversed RDS since the reversed depth from the first-order mechanism and the normal depth from the second-order mechanism conflict. Noting that the second-order process is stronger in central vision, they explained the finding(Zhaoping and Ackermann, 2018) that reversed depth is perceived peripherally but not centrally. Through their model, our congruent RDSs should evoke stronger first-order normal depth signal and weaker second-order normal depth signal, whereas our incongruent RDS should evoke weaker first-order normal depth sigal and stronger second order normal depth signal.

4.3.4 Feedback processes for depth perception

Our finding that the perceptual impact of the reversed depth signals depends on the viewing duration suggests that feedforward processes alone is insufficient. Feedback in the FFVW process helps us to make sense of the dependence on the temporal (viewing duration) and spatial (dot size) characters of the inputs, without evoking additional free parameters to adjust the relative weights between multiple phenomenological mechanisms to explain a diversity of data.

Our inferences about the top-down feedback processes in depth perception are consistent with that implied by neurophysiological data involving noisy RDSs and behaving monkeys(Nienborg and Cumming, 2009). Note, though, that depth noise in this monkey study was induced by temporal fluctuations of the disparity of a RDS surface, rather than using monocular noise as in our stimulus. The monocular noise dots in one eye can accidentally match monocular noise or even signals from the other eye. These accidental matches generate ghost depth signals, which were perceptually visible to our observers given a sufficiently long viewing duration. In this sense, our noisy RDS is analogous to the noisy motion signals used in coherent motion direction discrimination task in some monkey studies(Britten et al., 1996).

The choice of 0.02 seconds in the current study as an ultra-short presentation duration which would reduce the effect of the feedback verification was motivated by a $\sim 30 - 40$ ms latency between the feedforward and feedback components in visual cortical areas of monkeys. These components are identified by examining the temporal evolution of neural responses and their relationship with the visual inputs versus the task requirements(Chen et al., 2014, 2017; Yan et al., 2018) and by microstimulation studies in monkey visual cortex(Klink et al., 2017). In challenging visual discrimination of a brief visual target, a subsequent visual object, whose contour is close to that of the target, greatly impairs discrimination when presented 45 ms after the target's onset(Enns and Di Lollo, 1997). Assuming that this impairment, called object substitution, is due to a disruption of feedback verification by the subsequent visual input, its temporal character suggests a feedback latency similar to that by the monkey studies.

4.3.5 Binocular opponency processing

Processing of contrast-reversed RDSs by V1 and subsequent brain areas is another example of brain's computation with binocular opponency signals, or binocular differencing signals, defined as the difference between the left eye's input and the right eye's input. Up to the second-order, or pair-wise, correlation, these opponency signals form an independent information channel from the binocular summation channel via an efficient encoding of visual inputs(Li and Atick, 1994), and this encoding is manifested in V1's neural receptive fields. This opponency channel has been demonstrated recently by showing that visual adaptation to the opponency signal gives aftereffects in visual perception of motion(May et al., 2012), tilt(May and Zhaoping, 2016), and even faces(May and Zhaoping, 2019). Disparity tuning of V1 neurons comes from multiplexing the binocular summation and binocular opponency signals(Zhaoping, 2014). Thus, adapting to the binocular opponent signals also leads to aftereffects in perceived depth(Kingdom et al., 2020). Spatial contrast in this opponency signal can be very salient, attracting attention and gaze shifts(Zhaoping, 2008), and contrast-reversed RDSs can lead to a fast vergence response(Masson et al., 1997). The CPD hypothesis was partly motivated by the observation that this opponency signal is downplayed in visual perception more strongly in the central rather than the peripheral visual field(Zhaoping, 2017). The current study reveals additional ways to allow this opponency signal to contribute to perception, particularly in congruent RDSs.

4.4 Conclusion

In summary, hetero-pairs of dots across two eyes provide a rich opportunity to study feedforward and feedback processes in central vision. The current study uses them to test the Feedforward-Feedback-Verify-reWeight (FFVW) processes of visual inference. Although reversed depth signals evoked by hetero-pairs of dots in dense RDSs are typically invisible in central vision, they can impact depth perception in congruent and incongruent RDSs. They make depth surfaces in noisy RDSs more clearly perceived when they are congruent with the depth signals from homo-pairs of dots. However, when they are incongruent, they also make these depth surfaces less clearly perceived, but only when the RDSs are transiently presented to avoid the feedback veto in the FFVW process. Making the stimulus dots smaller increases the negative impact by the incongruent RDSs, consistent with the FFVW. The asymmetry between congruent and incongruent sensory inputs for perception reveals a constructive nonlinearity in the analysis-by-synthesis nature of the feedback component in the FFVW process. This constructive nonlinearity is akin to filling-in and visual completion when imperfect sensory signals are sufficiently adequate. The findings in this study should motivate future investigations to test FFVW further and reveal the underlying neural mechanisms and their perceptual consequences.

Acknowledgement: This work was supported by fundings from the Max Planck Society and University of Tübingen.

References

Anstis, S. (1970). Phi movement as a subtraction process. *Vision research*, 10(12):1411–IN5.

Aoki, S. C., Shiozaki, H. M., and Fujita, I. (2017). A relative frame of reference underlies reversed depth perception in anticorrelated random-dot stereograms. *Journal of Vision*, 17(12):17–17.

Asher, J. M. and Hibbard, P. B. (2018). First-and second-order contributions to depth perception in anti-correlated random dot stereograms. *Scientific reports*, 8(1):1–19.

Britten, K., Newsome, W., Shadlen, M., Celebrini, S., and Movshon, J. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, 13:87–100.

Carpenter, G. and Grossberg, S. (1987). Art 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26(23):4919–4930.

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., and Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron*, 82(3):682–694.

Chen, R., Wang, F., Liang, H., and Li, W. (2017). Synergistic processing of visual contours across cortical layers in V1 and V2. *Neuron*, 96(6):1388–1402.

Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L., and Blake, R. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception*, 24(1):33–47.

Crick, F. and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375(6527):121–3.

Cumming, B. G. and Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 389(6648):280–287.

Cumming, B. G., Shapiro, S. E., and Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception*, 27(11):1367–1377.

Doi, T. and Fujita, I. (2014). Cross-matching: a modified cross-correlation underlying threshold energy model and match-based depth perception. *Frontiers in computational neuroscience*, 8:127.

Doi, T., Takano, M., and Fujita, I. (2013). Temporal channels and disparity representations in stereoscopic depth perception. *Journal of Vision*, 13(13):26–26.

Doi, T., Tanabe, S., and Fujita, I. (2011). Matching and correlation computations in stereoscopic depth perception. *Journal of Vision*, 11(3):1.

Enns, J. T. and Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological science*, 8(2):135–139.

Fujita, I. and Doi, T. (2016). Weighted parallel contributions of binocular correlation and match signals to conscious perception of depth. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1697):20150257.

Helmholtz, H. v. (1925). *Physiological Optics (translated by J P C Southall)*. New York: Optical Society of America, Dover Press.

Henriksen, S., Read, J. C., and Cumming, B. G. (2016). Neurons in striate cortex signal disparity in half-matched random-dot stereograms. *Journal of Neuroscience*, 36(34):8967–8976.

Hibbard, P. B., Scott-Brown, K., Haigh, E., and Adrain, M. (2014). Depth perception not found in human observers for static or dynamic anti-correlated random dot stereograms. *PLoS One*, 9(1):e84087.

Julesz, B. (1971). Foundations of cyclopean perception. *University of Chicago Press*.

Kingdom, F. A., Yared, K.-C., Hibbard, P. B., and May, K. A. (2020). Stereoscopic depth adaptation from binocularly correlated versus anti-correlated noise: Test of an efficient coding theory of stereopsis. *Vision Research*, 166:60–71.

Klink, P. C., Dagnino, B., Gariel-Mathis, M.-A., and Roelfsema, P. R. (2017). Distinct feedforward and feedback effects of microstimulation in visual cortex reveal neural mechanisms of texture segregation. *Neuron*, 95(1):209–220.

Li, Z. and Atick, J. J. (1994). Efficient stereo coding in the multiscale representation. *Network: Computation in Neural Systems*, 5(2):157–174.

Lippert, J. and Wagner, H. (2001). A threshold explains modulation of neural responses to opposite-contrast stereograms. *Neuroreport*, 12(15):3205–3208.

MacKay, D. (1956). Towards an information flow model of human behavior. *British Journal of Psychology*, 47(1):30–43.

Masson, G., Busettini, C., and Miles, F. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, 389(6648):283–286.

May, K., Zhaoping, L., and Hibbard, P. (2012). Perceived direction of motion determined by adaptation to static binocular images. *Current Biology*, 22:28–32.

May, K. A. and Zhaoping, L. (2016). Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns. *Current Biology*, 26(12):1571–1576.

May, K. A. and Zhaoping, L. (2019). Face perception inherits low-level binocular adaptation. *Journal of vision*, 19(7):7–7.

Neri, P., Parker, A. J., and Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, 401(6754):695–698.

Nieder, A. and Wagner, H. (2001). Hierarchical processing of horizontal disparity information in the visual forebrain of behaving owls. *Journal of Neuroscience*, 21(12):4514–4522.

Nienborg, H. and Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 459(7243):89–92.

Ohzawa, I., DeAngelis, G., and Freeman, R. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972):1037–1041.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, 6(3):390–404.

Qian, N. and Mikaelian, S. (2000). Relationship between phase and energy methods for disparity computation. *Neural Computation*, 12(2):279–292.

Read, J. C. and Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision research*, 40(24):3345–3358.

Tanabe, S., Umeda, K., and Fujita, I. (2004). Rejection of false matches for binocular correspondence in macaque visual cortical area v4. *Journal of Neuroscience*, 24(37):8170–8180.

Tanabe, S., Yasuoka, S., and Fujita, I. (2008). Disparity-energy signals in perceived stereoscopic depth. *Journal of vision*, 8(3):22–22.

Tanaka, H. and Ohzawa, I. (2006). Neural basis for stereopsis from second-order contrast cues. *Journal of Neuroscience*, 26(16):4370–4382.

Wilcox, L. M. and Hess, R. F. (1996). Is the site of non-linear filtering in stereopsis before or after binocular combination? *Vision Research*, 36(3):391–399.

Yan, Y., Zhaoping, L., and Li, W. (2018). Bottom-up saliency and top-down learning in the primary visual cortex of monkeys. *Proceedings of the National Academy of Sciences*.

Yuille, A. and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308.

Zhaoping, L. (2008). Attention capture by eye of origin singletons even without awareness—a hallmark of a bottom-up saliency map in the primary visual cortex. *Journal of Vision*, 8(5):article 1.

Zhaoping, L. (2012). Gaze capture by eye-of-origin singletons: Interdependence with awareness. *Journal of Vision*, 12(2):article 17.

Zhaoping, L. (2014). Understanding vision: theory, models, and data. *Oxford University Press*.

Zhaoping, L. (2017). Feedback from higher to lower visual areas for visual recognition may be weaker in the periphery: Glimpses from the perception of brief dichoptic stimuli. *Vision Research*, 136:32–49.

Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology*, 58:1–10.

Zhaoping, L. (2020). The flip tilt illusion: Visible in peripheral vision as predicted by the central-peripheral dichotomy. *i-Perception*, 11.

Zhaoping, L. and Ackermann, J. (2018). Reversed depth in anticorrelated random-dot stereograms and the central-peripheral difference in visual inference. *Perception*, 47(5):531–539.

Zhaoping, L. and Jingling, L. (2008). Filling-in and suppression of visual perception from context: A Bayesian account of perceptual biases by contextual influences. *PLoS Computational Biology*, 4(2):e14.