Combining Sensory Information: Mandatory Fusion Within, but Not Between, Senses

J. M. Hillis,[1]*†‡ M. O. Ernst,[2]* M. S. Banks,[1,3] M. S. Landy[4]

Humans use multiple sources of sensory information to estimate environmental properties. For example, the eyes and hands both provide relevant information about an object's shape. The eyes estimate shape using binocular disparity, perspective projection, etc. The hands supply haptic shape information by means of tactile and proprioceptive cues. Combining information across cues can improve estimation of object properties but may come at a cost: loss of single-cue information. We report that single-cue information is indeed lost when cues from within the same sensory modality (disparity and texture gradients in vision) are combined, but not when different modalities (vision and haptics) are combined.

Sensory estimates of an environmental property can be represented by $\hat{S}_i = f_i(S)$ where S is the physical property being estimated, f is the operation the nervous system performs to derive the estimate, and $\hat{S}$ is the perceptual estimate. The subscripts refer to different sensory modalities (e.g., haptics and vision) or different cues within a modality (e.g., disparity and texture cues within vision). Sensory estimates are subject to two types of error: random measurement error and bias. Thus, estimates of the same object property from different cues usually differ. To reconcile the discrepancy, the nervous system must either combine estimates or choose one, thereby ignoring the other cues. Assuming that each single-cue estimate is unbiased (*1*) but corrupted by independent Gaussian noise, the statistically optimal strategy for cue combination is a weighted average (*2*, *3*)

$$\hat{S}_c = \sum_i w_i \hat{S}_i, \text{ where } w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \quad (1)$$

w_i is the weight given to the ith single-cue estimate, and σ_i^2 is that estimate's variance. Combining estimates by this maximum-likelihood estimation (MLE) rule yields the least variable estimate of S and thus more precise estimates of object properties (*4*–*6*).

To benefit from MLE (in the sense of reducing uncertainty), different cues to the same object property must be well correlated across objects. For example, if an object's size increases, visual and haptic signals both generally indicate the increase, so an organism would obtain the benefit of more precise estimates by using MLE. There is, however, a potential cost: Consider the situation in which there are two cues, S_1 and S_2. In this case, the MLE is

$$\hat{S}_c = w_1\hat{S}_1 + w_2\hat{S}_2 \quad (2)$$

There are combinations of S_1 and S_2, producible in the laboratory, for which $\hat{S}_c$ is, on average, constant. If $S_1 = S + \Delta S_1$ and $S_2 = S + \Delta S_2$, then $\hat{S}_c$ is constant, on average, for values of ΔS_1 and ΔS_2 satisfying

$$\Delta S_2 = -\frac{w_1}{w_2}\Delta S_1 \quad (3)$$

If the combined estimate (Eq. 2) were the only one available, the nervous system would be unable to discriminate the various stimulus combinations satisfying Eq. 3. Such physically distinct, but perceptually indiscriminable, stimuli would be metamers (*7*–*9*). If, in contrast, the nervous system retained the single-cue estimates, $\hat{S}_1$ and $\hat{S}_2$, the various combinations satisfying Eq. 3 would be discriminable from one another. An inability to discriminate stimulus combinations satisfying Eq. 3 would have little practical consequence because such combinations rarely occur in the natural environment. We can, however, generate such combinations in the laboratory and then look for the loss in discrimination capability predicted by MLE. Observing such a loss would mean that the nervous system combines information from different cues to form one estimate of the object property in question. (That is, it would mean that mandatory cue fusion occurs.)

In previous studies on cue combination, participants were asked to report perceived location (*6*), shape (*10*), slant (*11*), or distance (*12*). These studies tell us that multiple cues are used for the associated judgment, but not whether participants lose access to single-cue estimates. To remedy this, we used an oddity task. Participants identified, among three stimuli, the stimulus that differed from the other two on any dimension. Because participants were free to use any difference, not just the one imposed by the experimenter, the task provided a true test for the existence of cue fusion.

To understand our experimental design, consider Fig. 1. Figure 1A shows surface slant specified by disparity (abscissa, S_1) and texture (ordinate, S_2). At the origin, disparity- and texture-defined slants are frontoparallel (0,0). The blue diagonal represents other cases in which the disparity and texture slants are equal. The enlargement in the middle shows one case in which the origin is a stimulus whose disparity- and texture-specified slants are nonzero, but equal (S,S). We call this the standard stimulus, an example of which is shown to the right of the enlarged graph [supporting online material (SOM) Text]. The standard contained consistent cues, so it was always on the blue diagonal. The abscissa in the enlargement represents changes in S_1 and the ordinate changes in S_2. Stimuli resulting from those changes are comparison stimuli (see corresponding examples, also to the right).

How different must the comparison be from the standard for a person to distinguish the two stimuli? By distinguish, we mean the participant can reliably pick out the "odd" stimulus when presented two examples of the standard and one of the comparison (or vice versa). There are at least two strategies. (i) With single-cue estimators, estimates could be made independently of each other (without combination). Once the difference between either estimate, ΔS_1 or ΔS_2, reached its own discrimination threshold, $\pm T_1$ or $\pm T_2$, the participant would be able to identify the odd stimulus. The predicted set of thresholds would, therefore, be horizontal and vertical lines (Fig. 1B, left; red lines) whose distance from the origin was determined by the discrimination threshold of the single-cue estimators (*13*). Comparison stimuli within the rectangle would be indistinguishable from the standard. Performance would be no better in one quadrant than in any other. (ii) With a mandatory combined estimator, only the combined estimate (Eq. 2) is used. Discrimination threshold would be determined only by the difference in the combined estimates for the standard and comparison. Assuming that the variances of the single-cue estimates are equal to the variances used to plot the single-cue estimation prediction (Fig. 1B, left) and that the MLE rule is used, the pre-

[1]Vision Science Program, School of Optometry, University of California, Berkeley, CA 94720–2020, USA. [2]Max-Planck Institute for Biological Cybernetics, Spemannstraβe 38, 72076, Tübingen, Germany. [3]Department of Psychology, University of California, Berkeley, CA 94720–1650, USA. [4]Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA.

*These authors contributed equally to this work.
†Present address: University of Pennsylvania, Department of Psychology, 3815 Walnut Street, Philadelphia, PA, 19104, USA.
‡To whom correspondence should be addressed. E-mail: jmhillis@CATTELL.psych.upenn.edu

dicted thresholds for the combined estimation strategy are the green diagonal lines (Fig. 1B, middle).

To determine how cues are used in discriminating object properties, we conducted two discrimination experiments. One experiment involved cues in different modalities (vision and touch) and the other involved cues in the same modality (disparity and texture within vision). We measured size discrimination in the intermodal experiment and slant discrimination in the within-modal experiment (*14*).

Mandatory combination, i.e., combining cues and losing access to individual estimates, is more likely in the within-modal (disparity-texture) case than the intermodal (visual-haptic) case. In the intermodal case, there are circumstances in the natural environment when one touches one object while looking at another. In this case, it would be disadvantageous to combine haptic and visual estimates. For the within-modal case, the two cues are always spatially coincident, so MLE combination might be implemented all the time.

In the visual-haptic experiment (*15*), we measured single-cue (vision alone and haptics alone) and intermodal discrimination (visual-haptic). The stimulus was a horizontal bar raised 30 mm above a plane; the plane and the bar's front surface were perpendicular to the line of sight (Fig. 1A, bottom). The visual stimulus was a random dot stereogram simulating the background plane and bar (fig. S1). On half of the trials, visual "noise" was added into the stereogram (i.e., random displacements parallel to the line of sight were added to each dot). The haptic stimulus was generated using two PHANToM force-feedback devices (Fig. 1C), one each for the index finger and thumb. Participants viewed the bar binocularly and/or grasped it with the index finger and thumb to estimate its height. In the visual-haptic trials, the visually specified bar did not appear until the bar was touched by both fingers simultaneously. The visual and haptic stimuli disappeared after 1 s.

Each trial consisted of the sequential presentation of three bars. The standard was presented twice and the comparison, once. The participant indicated the one that was different from the other two on any basis. No feedback was given. For both single-cue and intermodal conditions, the standard bar height was 55 mm and the comparison was selected randomly from a set of predetermined heights. In the intermodal condition, the visual (v) and haptic (h) comparison height was specified by a fixed ratio ($\Delta S_h/\Delta S_v$). Each ratio corresponds to a direction in the stimulus space (Fig. 1A). We tested six directions.

The upper right and lower left panels of Fig. 2A show error rate (1 – proportion correct) for one participant in the vision-alone and haptic-alone conditions, respectively. The solid curves are best-fitting Gaussian curves to the error-rate data. Threshold was defined as one standard deviation from the mean; those values are indicated by the red dashed lines.

The lower right panel of Fig. 2A shows intermodal data for the same participant. These data show that a combined estimator is used in visual-haptic judgments. As expected from use of a combined estimator, thresholds were lower in quadrants 1 and 3 (where cues are consistent) than in quadrants 2 and 4 (cues inconsistent). The data also show that single-cue estimators are used: discrimination thresholds were not consistently higher in quadrants 2 and 4 than predicted by use of single-cue estimators (Fig. 1B, right) (*16*).

Because it seems more likely that one would observe evidence for mandatory fusion when the cues are combined within one modality, we conducted an experiment in which we manipulated two visual cues to slant: disparity and texture.

We used a custom stereoscope to present the displays (*11*) (fig. S2). The stimuli were planes slanted about a vertical axis. In most cases, they were viewed binocularly. The surfaces were textured with Voronoi patterns (*17*, *18*) generated from a jittered grid of dots (different for each presentation).

Each trial consisted of three sequential presentations and participants indicated the one that was different on any basis. No feedback was given. Three of six participants were presented with 40 different ratios of $\Delta S_t/\Delta S_d$ (i.e., 40 directions in Fig. 1A; t indicates texture and d indicates disparity), and the other three were presented with a subset. At least 10 ratios were presented in each session. The difference between the standard and the comparison along each tested direction (i.e., each $\Delta S_t/\Delta S_d$) was varied according to an adaptive staircase, a different randomly interleaved staircase for each direc-

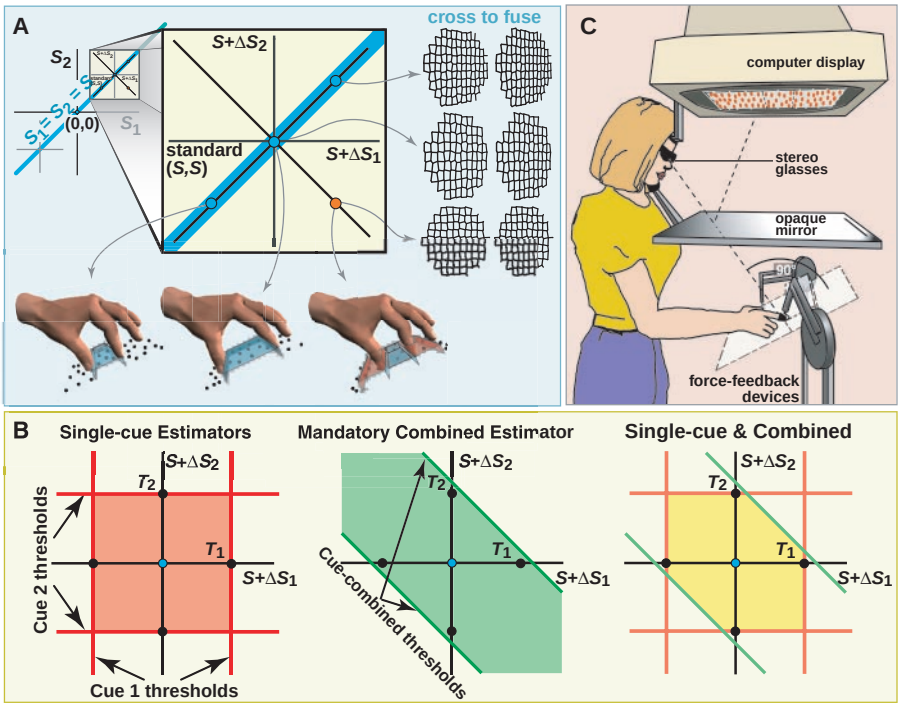

Fig. 1. Stimuli and predictions. (**A**) The abscissa represents changes in S_1 (disparity-specified height for the intermodal case and disparity-specified slant for the within-modal case). The ordinate represents changes in S_2 (haptically-specified height for the intermodal case and texture-specified slant in the within-modal case). The stereograms (SOM Text) on the right illustrate the standard, within-modal stimulus (middle); a comparison with disparity and texture specifying a larger slant than the standard (top); and another comparison with disparity specifying a larger slant and texture specifying a lesser slant relative to the standard (bottom). The icons below the plot represent the corresponding interpretations for the intermodal experiment. (**B**) (Left) Predicted discrimination thresholds if participants use independent single-cue estimates. Comparison stimuli within the rectangle would be indistinguishable from the standard. (Middle) Predicted thresholds if participants have access to the combined estimate only. Comparison stimuli between the two lines would be indistinguishable from the standard. (Right) Predicted discrimination thresholds if participants have access to the single-cue and combined estimates. (**C**) Visual-haptic apparatus. Participants viewed binocularly the reflection of the visual stimulus presented on a computer display. Crystal Eyes (StereoGraphics, San Rafael, CA) liquid-crystal shutter glasses were used to present binocular disparity. A head and chin rest limited head movements. The right hand was beneath the mirror and could not be seen.

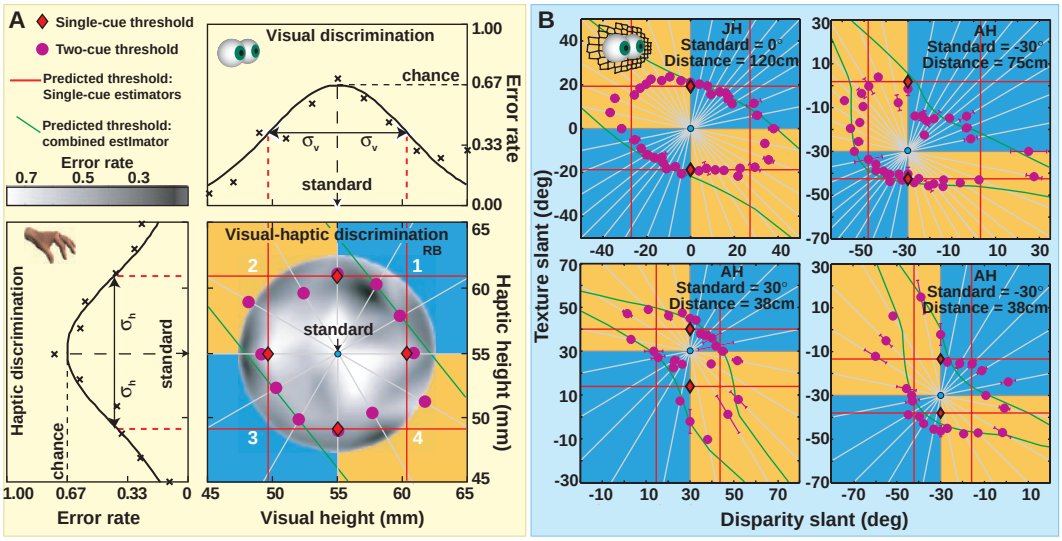

Fig. 2. Discrimination thresholds in the inter- and within-modal experiments. (**A**) Results for one participant in the intermodal (visual-haptic) experiment (see fig. S4 for results from other participants). In the bottom right panel, the red lines represent predicted single-cue estimator thresholds and the green lines represent predicted combined estimator thresholds. The magenta points are discrimination thresholds for one participant in the visual-haptic task. Lightness within the central gray circle represents the participant's error rate. Quadrants (numbered in white) 1 and 3, where combined estimation predicts improvement relative to single-cue estimation, are colored blue. Quadrants 2 and 4, where combined estimation predicts deterioration relative to single-cue estimation, are colored orange. (**B**) Four examples of within-modal (disparity-texture) results plotted in the same format. The red diamonds represent thresholds for the texture estimator (separate monocular experiment). The red lines represent the predicted thresholds from the single-cue estimator. The green curves from the combined estimator (Eq. 2) curve because the weights assigned to an optimal combined estimator vary with texture-specified slant (*17*) (SOM Text). The magenta points represent thresholds in the various directions ($\Delta S_t/\Delta S_d$) tested in the discrimination experiment.

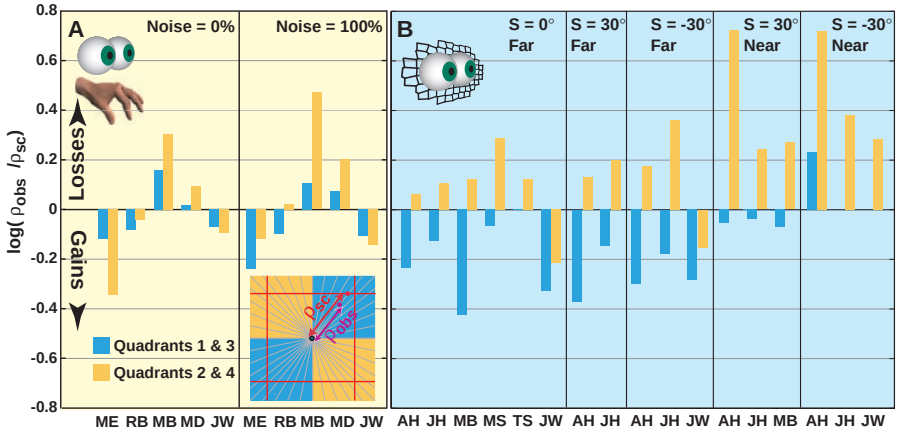

Fig. 3. Asymmetry of threshold data in the inter- and within-modal experiments. Summary results for the intermodal (**A**) and within-modal cases (**B**). Blue bars represent results from quadrants 1 and 3, where cues are consistent and combined estimation would yield improvement relative to single-cue estimation. Orange bars represent results from quadrants 2 and 4, where cues are inconsistent and combined estimation predicts poorer performance relative to single-cue estimation. The bar heights are the mean of the log of ρ_{obs}/ρ_{sc}. Negative values indicate improvement relative to single-cue estimation, and positive values indicate deterioration.

tion. We collected data at various viewing distances and base slants.

Figure 2B shows some results for two participants. Data followed the curves predicted from optimal combination (SOM Text; fig. S3), which means that texture and disparity weights are set dynamically, trial-by-trial, according to the relative variances of the texture and disparity estimators (*17*). Some thresholds actually fell outside the rectangle defined by single-cue thresholds (red lines) (*19*). In these cases, the participant failed to discriminate, on any basis, stimuli whose constituents were themselves discriminable. The standard and comparison were metamers (*7–9*); i.e., we observed mandatory fusion within vision.

Our main interest is to determine whether gains and losses associated with using a combined estimator occur, so we compared all thresholds in quadrants 1 and 3 to those in quadrants 2 and 4. For each direction in a quadrant, there is a threshold predicted from use of single-cue estimators (ρ_{sc}, the distance from the origin to the nearest red line; Fig. 3A, lower right) and an observed threshold (ρ_{obs}, distance from the origin to the data point). We can express gains and losses relative to single-cue estimation as the log ratio of observed and predicted single-cue thresholds: log (ρ_{obs}/ρ_{sc}). For all conditions and participants, we computed the average log ratio for the cues-consistent quadrants (1 and 3) and for the cues-inconsistent quadrants (2 and 4). Results from the visual-haptic and disparity-texture experiments are presented in Fig. 3, A and B, respectively. Bar color represents the quadrants from which the data were obtained. In nearly all cases, the log ratios in the cues-consistent quadrants were less than in the cues-inconsistent ones. In those cases, the discrimination data were asymmetric, as expected from use of a combined estimator (i.e., thresholds were lower when the changes specified by the two cues were in the same direction than when they were in opposite directions). However, only the within-modal data consistently indicated losses (log (ρ_{obs}/ρ_{sc}) > 0), sometimes rather large ones, when the texture and disparity cues specified slant changes in opposite directions.

The absence of large threshold elevations in the inconsistent-cues direction for the intermodal case is not surprising. Haptic and visual signals for size do not always come from the same object (e.g., touching one object while looking at another). Mandatory combination of haptic and visual cues would be misleading in such cases. The situation is different in the within-modal case: texture and disparity cues at the same retinal location almost always come from the same object. Thus, mandatory cue combination would be beneficial if errors in the texture and disparity estimates were a more likely cause of discrepancy than actual signal differences. For this reason, there would be evolutionary or developmental pressure to rely on the combined estimate instead of the single-cue estimates. This deferral to the combined estimate is evident in the deterioration of performance in the within-modal data (Figs. 2B and 3B) in the cues-inconsistent quadrants (2 and 4); participants would have performed better if they used single cues. To highlight this

point, we measured thresholds with only one cue present (monocular texture) and with both cues present (texture and disparity) in sequential blocks of trials. For the cues-inconsistent quadrants, thresholds were lower in the monocular condition than when both disparity and texture were available. The opposite was true in the cues-consistent quadrants. This result illustrates both the benefits (better discrimination when the cues specify changes in the same direction) and the costs (the loss of single-cue information associated with cue combination).

Our data provide a clear demonstration of depth-cue fusion: shape information from texture and disparity cues is combined to form a single, fused percept such that some discriminations that could be made from single-cue estimates are not made (*19*). We also have evidence for a single, fused percept for shape information from haptics and vision, but in this intermodal case information from single-cue estimates is not lost.

References and Notes

1. Our interest is in the manner in which the nervous system resolves discrepancies at any given moment between sensory measurements. We assume the sensory systems under examination are well calibrated, so their signals will, on average, agree with one another. They will, however, disagree from one measurement to the next due to random measurement error.
2. J. J. Clark, A. L. Yuille, *Data Fusion for Sensory Information Systems* (Kluwer Academic, Boston, MA, 1990).
3. M. S. Landy, L. T. Maloney, E. B. Johnston, M. J. Young, *Vision Res*. **35**, 389 (1995).
4. Z. Ghahramani, D. M. Wolpert, M. I. Jordan, in *Self-Organization, Computational Maps, and Motor Control*, P. G. Morasso, V. Sanguineti, Eds. (Elsevier, Amsterdam, 1997) pp. 117–147.
5. M. O. Ernst, M. S. Banks, *Nature* **451**, 429 (2002).
6. M. S. Landy, H. Kojima, *J. Opt. Soc. Am.* **18**, 2307 (2001).
7. Metamers are composite stimuli that cannot be discriminated even though their constituents can be. The classic example is the inability to discriminate a yellow light consisting of one wavelength from another yellow light consisting of red added to green.
8. W. Richards, *Sens. Processes* **3**, 207 (1979).
9. B. T. Backus, in *Advances in Neural Information Processing Systems*, vol. 14, T. G. Dietterich, S. Becker, Z. Ghahramani, Eds. (MIT, Cambridge, MA, 2002) part VII, chap. 1.
10. E. B. Johnston, B. G. Cumming, M. S. Landy, *Vision Res.* **34**, 2259 (1994).
11. B. T. Backus, M. S. Banks, R. van Ee, J. A. Crowell, *Vision Res*. **39**, 1143 (1999).
12. E. Brenner, W. J. M. van Damme, *Vision Res.* **39**, 975 (1999).
13. With two independent estimators, there are two chances for discriminating the odd stimulus. For each value of the comparison, each estimator has a likelihood of discriminating the comparison from the standard. Complete predictions for independent estimators would therefore include probability summation: P(estimator 1 discriminates odd) + P(estimator 2 discriminates odd) – P(both discriminate odd). Predicted thresholds that include probability summation would be a rectangle like the one in Fig. 1B, but with rounded corners. Inclusion of probability summation does not affect our interpretation of the data, because probability summation would not produce the observed asymmetries between thresholds in the four quadrants (Figs. 2 and 3).
14. From among the large set of possible tasks and cues, we have examined a small subset. We chose tasks that were natural for participants to perform and pairs of cues that were approximately equally reliable to better show the gains and losses predicted by MLE.
15. Material and methods are available as supporting online material on *Science* Online.
16. The participants' phenomenology was instructive. They reported using a difference in perceived size when the comparison stimulus was in the cues-consistent quadrants (1 and 3). This percept is well modeled by the equations for combined estimation (Eqs. 1 and 2). Participants' reports were less consistent with stimuli in the cues-inconsistent quadrants (2 and 4). Sometimes they used a difference in perceived size, but frequently they noticed the conflict between the visually and haptically specified sizes and used the perceived conflict to make the oddity discrimination. The phenomenology is consistent with the hypothesis that participants used three estimators in performing the oddity discrimination: two single-cue estimators and a combined estimator.
17. D. C. Knill, *Vision Res.* **38**, 1683 (1998).
18. M. de Berg, M. van Kreveld, M. Overmars, O. Schwarzkopf, *Computational Geometry: Algorithms and Applications* (Springer-Verlag, New York, ed. 2, 2000).
19. The participants' phenomenology was informative. In some trials, they perceived a difference in slant (quadrants 1 and 3). This percept is well modeled by the equations for combined estimation (Eqs. 1 and 2). In other trials, they perceived a difference in the shape of the surface texture. This occurred in the cues-inconsistent quadrants (particularly in the direction for which S_c is constant, Eq. 3). For example, in the lower right panel of Fig. 2B, the standard's slant was –30 (i.e., left side, far), so the retinal images had smaller and more foreshortened texture on the left. For the four comparison stimuli in quadrant 2, the disparity- and texture-specified slants differed in opposite directions from the standard stimulus. $\hat{S}_c$ was approximately –30 for these comparison stimuli, but the surface texture was rendered such that the retinal images were nearly equally large (and equally foreshortened) on the left and right. To be consistent with the perceived left-far slant, the surface texture would have to be nonhomogeneous (larger cells on the left and smaller on the right). The odd stimulus was detected by perceiving this nonhomogeneity. Participants were, in this case, using shape-constancy mechanisms (which allow one to determine the shape of markings on a slanted surface). We believe, therefore, that participants did not have access to more than one percept (e.g., $\hat{S}_c$, $\hat{S}_t$ and $\hat{S}_d$); the slant cues were truly fused. They made the discrimination in the cues-inconsistent quadrants by use of another calculation—shape constancy—that allowed them to perceive the objective shape of the texture on the surface. If this hypothesis is correct, metamers (cases in which discrimination is poorer than predicted by single-cue estimates) occurred when the stimulus made the shape-constancy judgment difficult.

20. We thank S. Watt, S. Gepshtein, and B. Backus for comments. This work was supported by the Max-Planck Society and by research grants from Air Force Office of Scientific Research (F49620-98), NIH (EY08266 and EY12851), NSF (DBS-9309820) and by an equipment grant from Silicon Graphics.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5598/1627/DC1
Materials and Methods
SOM Text
Figs. S1 to S4

24 June 2002; accepted 25 September 2002

A Critical Role for IL-21 in Regulating Immunoglobulin Production

Katsutoshi Ozaki,[1]* Rosanne Spolski,[1] Carl G. Feng,[2] Chen-Feng Qi,[3] Jun Cheng,[4] Alan Sher,[2] Herbert C. Morse III,[3] Chengyu Liu,[5] Pamela L. Schwartzberg,[4] Warren J. Leonard[1]†

The cytokine interleukin-21 (IL-21) is closely related to IL-2 and IL-15, and their receptors all share the common cytokine receptor γ chain, γ_c, which is mutated in humans with X-linked severe combined immunodeficiency disease (XSCID). We demonstrate that, although mice deficient in the receptor for IL-21 (IL-21R) have normal lymphoid development, after immunization, these animals have higher production of the immunoglobulin IgE, but lower IgG1, than wild-type animals. Mice lacking both IL-4 and IL-21R exhibited a significantly more pronounced phenotype, with dysgammaglobulinemia, characterized primarily by a severely impaired IgG response. Thus, IL-21 has a significant influence on the regulation of B cell function in vivo and cooperates with IL-4. This suggests that these γ_c-dependent cytokines may be those whose inactivation is primarily responsible for the B cell defect in humans with XSCID.

The receptor for the lymphoid-specific cytokine IL-21 is expressed on T, B, and NK cells (*1*, *2*). IL-21 was initially reported to have a costimulatory T cell proliferative effect, to augment NK cell expansion and differentiation, and to augment B cell proliferation in response to CD40-specific antibodies, but to inhibit proliferation of B cells stimulated with the combination of IL-4 and IgM-specific antibodies (*2*). Subsequently, IL-21 was reported not to be required for NK cell development or expansion from murine splenocytes and to oppose certain actions of IL-15 on activated NK cells (*3*). The receptor for IL-21 contains IL-21R (*1*, *2*) and also shares the common cytokine receptor γ chain (γ_c) with IL-2, IL-4, IL-7, IL-9, and IL-15 (*4*, *5*). Mutations in γ_c result in XSCID (*4*, *6*), a disease characterized by an absence of T and NK cells and nonfunctional B cells (*7*). Although the absence of T and NK cells can be explained by defective responses to IL-7 (*8*–*10*) and IL-15 (*11*–*13*), no cytokine has been linked to the B cell defect. To determine