

Separability and Commonality of Auditory and Visual Bistable Perception

Hirohito M. Kondo¹, Norimichi Kitagawa¹, Miho S. Kitamura^{1,2}, Ai Koizumi^{1,3}, Michio Nomura^{1,4,7} and Makio Kashino^{1,5,6}

¹NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa 243-0198, Japan, ²Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan, ³Department of Psychology, The University of Tokyo, Tokyo 113-0033, Japan, ⁴Division of Human Sciences, Graduate School of Integrated Arts and Sciences, Hiroshima University, Higashi-Hiroshima, Hiroshima 739-8521, Japan, ⁵ERATO Shimojo Implicit Brain Function Project, Japan Science and Technology Agency, Atsugi, Kanagawa 243-0198, Japan and, ⁶Department of Information Processing, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan

⁷Current address: Department of Cognitive Psychology in Education, Graduate School of Education, Kyoto University, Kyoto 606-8501, Japan

Address correspondence to Hirohito M. Kondo, NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. Email: kondo.hirohito@lab.ntt.co.jp.

It is unclear what neural processes induce individual differences in perceptual organization in different modalities. To examine this issue, the present study used different forms of bistable perception: auditory streaming, verbal transformations, visual plaids, and reversible figures. We performed factor analyses on the number of perceptual switches in the tasks. A 3-factor model provided a better fit to the data than the other possible models. These factors, namely the "auditory," "shape," and "motion" factors, were separable but correlated with each other. We compared the number of perceptual switches among genotype groups to identify the effects of neurotransmitter functions on the factors. We focused on polymorphisms of catechol-O-methyltransferase (COMT) Val¹⁵⁸Met and serotonin 2A receptor (HTR2A) -1438G/A genes, which are involved in the modulation of dopamine and serotonin, respectively. The number of perceptual switches in auditory streaming and verbal transformations differed among COMT genotype groups, whereas that in reversible figures differed among HTR2A genotype groups. The results indicate that the auditory and shape factors reflect the functions of the dopamine and serotonin systems, respectively. Our findings suggest that the formation and selection of percepts involve neural processes in cortical and subcortical areas.

Keywords: awareness, brainstem, consciousness, human, illusion

Introduction

We perceive the world as stable, although sensory inputs are often ambiguous due to spatial and temporal occluders. This raises an important question regarding how stable percepts are formed in the brain. Bistable perception phenomena provide us with clues enabling us to investigate that issue because constant physical stimulation leads to spontaneous switching between different stable percepts. Although in the past, the formation and selection of percepts have been investigated with binocular rivalry, reversible figures, and visual plaids in the visual domain (Kleinschmidt et al. 1998; Tong et al. 1998; Castelo-Branco et al. 2002; Haynes et al. 2005; Wunderlich et al. 2005), more recently, they have been studied with auditory streaming and verbal transformations in the auditory domain (Gutschalk et al. 2005; Kondo and Kashino 2007, 2009; Schadwinkel and Gutschalk 2011). However, individual variation in perceptual switching has been overlooked in favor of averaging differences to focus on stimulus-specific commonalities.

An early study found a wide range of individual differences in the rate of binocular rivalry (Pettigrew and Miller 1998), where monocular images are presented to different eyes. The authors also pointed out that the rate of binocular rivalry is slow in patients with bipolar disorder, which is strongly heritable. This suggests that genetic factors influence the formation and selection of visual percepts. A large-sample twin heritability study has demonstrated that an approximately 50% variance in binocular rivalry rate is accounted for by additive genetic factors (Miller et al. 2010). In addition, a recent twin study confirmed that genetic factors affect the switching rate of the reversible figure as well as that of binocular rivalry (Shannon et al. 2011). These findings indicate that there is a substantial genetic contribution to bistable perception, particularly in the visual domain. However, it is unclear what neural processes are involved in individual differences in bistable perception and whether auditory bistability is functionally linked with visual bistability.

The present study elucidated the above issues using different forms of ambiguous stimuli. Perceptual switches in auditory streaming and verbal transformations are caused by prolonged listening to a sound sequence consisting of a triplet tone (van Noorden 1975; Bregman 1990) and word (Warren and Gregory 1958; Warren 1961). Perceptual switches in visual plaids and reversible figures are produced by observing moving gratings (Wallach 1935; Adelson and Movshon 1982; Hupé and Rubin 2003) and static figures (Long and Toppino 2004). There is still some controversy as to whether spontaneous perceptual switching is modulated by distributed processes within the sensory cortices (Pressnitzer and Hupé 2006) or a central oscillator within the subcortical areas (Pettigrew and Miller 1998). The present study employed 2 different approaches to clarify the relationship between auditory and visual bistability. First, we performed factor analyses of the number of the perceptual switches in the tasks and compared the fit indices of 1-factor and multifactor models. A factor analysis estimates the degree to which the variances of the observed variables can be explained by a small number of latent variables called factors. Thus, the analysis allows us to specify the underlying structure among observed and latent variables: The observed variables are modeled as linear combinations of the common factors and error terms. If perceptual switching in different modalities is governed by a single rhythm generator, the 1-factor model should provide a better fit to the data. Conversely, if different forms of bistable perception are implemented in

several brain modules, the multifactor model should fit the data. However, a factor analysis provides a heuristic interpretation of the results and cannot identify the functional linkage between the factors and neural processes.

Second, we used a genotype group comparison to examine which neurotransmitter functions are associated with the factors. Previous studies have argued that the timing of perceptual switching is modulated by the autonomic nervous system via noradrenaline (Einhäuser et al. 2008) and by drugs affecting the functions of the serotonin receptors (Carter et al. 2005, 2007; Nagamine et al. 2008). Thus, the dopamine and serotonin systems may be involved in the underlying neural processes of bistable perception. We focused on the functional polymorphisms of the catechol-*O*-methyltransferase (COMT) Val¹⁵⁸Met and serotonin 2A receptor (HTR2A) -1438G/A genes. COMT regulates the metabolism of catecholamines, such as dopamine and noradrenaline (Lotta et al. 1995), whereas the HTR2A gene is located in the promoter region and its activity increases cortical activation via glutamatergic excitatory post-synaptic potentials (Aghajanian and Marek 1997; Jakab and Goldman-Rakic 1998). Performance tends to be better for methionine (Met) allele carriers than for valine (Val) allele carriers in the Wisconsin Card Sorting Test (Egan et al. 2001) and the *n*-back task (Goldberg et al. 2003). Although the genetic effects of COMT on perception are unclear, a recent electroencephalographic (EEG) study showed that the amplitude of the N100 component was smaller for Met/Met individuals than for Val/Met and Val/Val individuals during an auditory task, arguing that the COMT genotype is associated with the poor sensory gating of auditory stimuli (Majic et al. 2011). The administration of psilocybin, namely an agonist of HTR1A and HTR2A, suppresses the rate of binocular rivalry (Carter et al. 2005). These findings suggest the possibility that bistable perception is modulated by the neurotransmitter functions of the brainstem nuclei (Pettigrew and Miller 1998; Carter and Pettigrew 2003; Sheppard and Pettigrew 2006). Thus, we hypothesize that common factors derived from bistable perception phenomena reflect the underlying processes of the dopamine and serotonin systems.

Materials and Methods

Participants

One hundred college students participated in the experiment. They were right-handed Japanese people with normal or corrected-to-normal vision and with normal hearing. None had any history of neurological or psychiatric illness. All participants gave written informed consent, which was approved by the ethics committee of NTT Communication Science Laboratories. Since 8 participants exhibited a large number of perceptual switches ($n > 150$), particularly in the verbal transformation task, we removed their data to maintain the normality of the distribution. As a result, we analyzed data from 92 participants (47 males and 45 females; mean age 27.3 years, range 20–34 years). We also performed analyses using data obtained from all the participants, but the results were essentially the same. The results for the 92 participants are reported below.

Stimuli and Task Procedures

All the participants performed 4 tasks, each of which had 2 conditions. The order of the conditions (five 90-s trials for each) was counter-balanced across the participants. The participants pressed a button on a keyboard to report their perceptions. Stimulus presentation and response collection were managed using the Psychophysics Toolbox running under MATLAB on a MacBook computer.

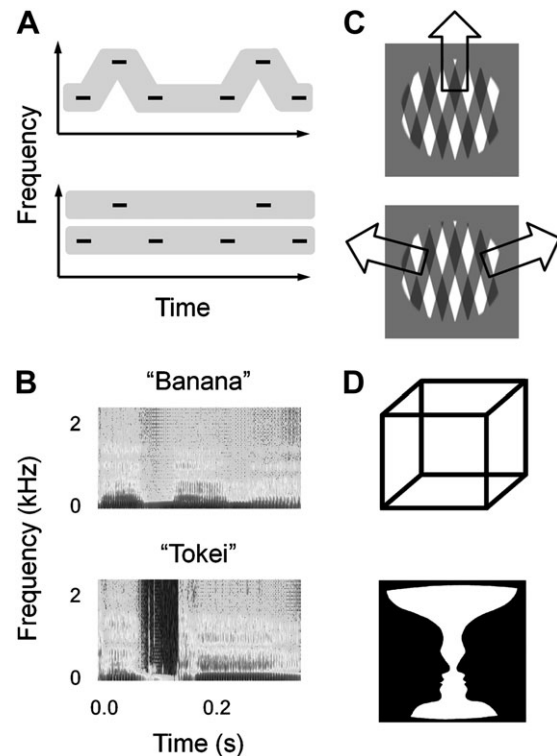


Figure 1. Schematic representation of auditory and visual stimuli. (A) A sound sequence of a triplet tone was presented in the auditory streaming task. The auditory stimulus produces perceptual switches between one stream with a galloping rhythm and 2 streams with isochronous rhythms. (B) The spectrograms indicate the time–frequency distribution of energy in the stimuli of the verbal transformation task. The constant stimulation of a stimulus word changes in the verbal form (e.g., “banana”) to another form (e.g., “nappa”). (C) Two rectangular gratings were moving in the visual plaid task. The visual stimulus induces perceptual switches between upward grouped and sideward split motion. (D) The Necker cube and Rubin vase were used in the reversible figure task. The Necker cube can be interpreted as having the lower left or upper right faces in front. Perceptual interpretation flips between black faces and a white vase in the Rubin vase.

In the auditory streaming task, the stimuli consisted of 225 repetitions of a triplet tone that comprised high (H) and low (L) tones with silent intervals (Fig. 1A). The duration of each tone was 40 ms, which included rising and falling cosine ramps of 10 ms. The stimulus onset asynchrony between the L tones was 200 ms, whereas that between the H tones was 400 ms. There were 2 frequency differences (Δf) between the H and L tones: H = 1069 Hz and L = 937 Hz at $\Delta f = 2$ semitones; H = 1213 Hz and L = 823 Hz at $\Delta f = 6$ semitones. In the verbal transformation task, the stimuli were 265 repetitions of the word “banana” or “tokei,” spoken by a female native speaker of Japanese (Fig. 1B). The Japanese word “tokei” means “clock” in English. The duration of each word was 340 ms without gaps. The auditory stimuli were presented at 65 dB SPL diotically through headphones (HDA 200; Sennheiser, Germany). Participants were instructed to close their eyes and listen to a sound sequence passively. They judged whether they had perceived a single coherent stream or 2 distinct streams in the auditory streams, and they indicated a perceptual switch from one verbal form to another in verbal transformations.

In the visual plaid task, the stimuli consisted of 2 rectangular-wave gratings: velocity = $1.25^\circ/\text{s}$; spatial frequency = $0.5 \text{ cycle}/^\circ$; duty cycle = 0.5 (Fig. 1C). The gratings were moving in directions ($\Delta\theta$) 100° and 150° apart. The stimuli were presented through a circular aperture on a gray background: viewing distance = 45 cm; visual angle = 5° . In the reversible figure task, a Necker cube and a Rubin vase were used as the stimuli (Fig. 1D). The participants judged whether they had perceived an upward grouped or sideward split motion in visual plaids, and they indicated their visual perception (e.g., black faces or white vase) as

regards the reversible figures. A small point for fixation appeared in the center of the visual stimuli.

SNP Genotyping

A single-nucleotide polymorphism (SNP) is a DNA sequence variation. SNPs commonly have 2 alleles, which are represented as combinations of a single nucleotide (e.g., C and C, C and T, or T and T). The form of the alleles (i.e., homozygous or heterozygous) differs among individuals, thus producing genetic variations within a population.

In this study, saliva samples were collected from all the participants and genomic DNA was extracted from the samples. G-to-A missense mutation results in Val to Met at codon 158 of the COMT gene, whereas a G/A polymorphism of the HTR2A gene is observed in the promoter region. The polymorphisms of the COMT Val¹⁵⁸Met (dbSNP accession: rs4680) and HTR2A -1438G/A (rs6311) genes were genotyped. We performed DNA amplification using a GeneAmp PCR System 9700 (Applied Biosystems, Tokyo, Japan). The primer pairs for the COMT were 5'-CACCTGTGCTCACCTCTCCT-3' and 5'-GGGTTTTTCAGTGAACGTGGT-3'; those for the HTR2A were 5'-AACCAACTTATTTCCTACCAC-3' and 5'-TAAGCTGCAAGGTAGCAACAG-3'. The reaction mixture was accommodated in a 20- μ L vessel containing sample DNA, 5 μ M of each primer, 2 mM of each dNTP, 10 \times polymerase chain reaction buffer, and AmpliTaq Gold (5 U/ μ L). Initial denaturation at 94 $^{\circ}$ C for 3 min was followed by 30 cycles of denaturation at 94 $^{\circ}$ C for 15 s, primer annealing at 60 $^{\circ}$ C for 30 s, and primer extension at 68 $^{\circ}$ C for 60 s. We determined polymorphisms using an Applied Biosystems 3730 DNA Analyzer with a BigDye Terminator Cycle Sequencing Kit. The HTR2A polymorphism of one participant could not be identified.

To compare the number of perceptual switches among genotype groups, we categorized the participants into 3 groups on the basis of their alleles: 18 Met/Met, 42 Val/Met, and 32 Val/Val individuals for the COMT polymorphism; 24 A/A, 46 G/A, and 21 G/G individuals for the HTR2A polymorphism. The COMT and HTR2A genotypes were distributed according to the Hardy-Weinberg equilibrium: $\chi^2_{(1)} = 0.39$ and 0.01, not significant (n.s.). The characteristics of the 3 groups did not differ in terms of gender and mean age: $\chi^2_{(2)} = 2.26$ and 1.10, n.s.; $F_{2,89} = 0.37$ and 0.25, n.s. Thus, we minimized the potential confounding factors leading to the contribution of the genetic variations to bistable perception.

Data Analyses

We computed the durations of alternating percepts (i.e., percept durations) for each trial and then obtained the number of perceptual switches for each individual. For each task, we determined that the data (sample size, $N = 92$) followed a normal distribution using the Kolmogorov-Smirnov test: $Z < 1.23$, $P > 0.10$. We performed exploratory factor and confirmatory factor analyses with SPSS 11.5J and Amos 4.0J.

In the exploratory factor analysis, we computed the Kaiser-Meyer-Olkin (KMO) statistic to assess sampling adequacy. The KMO overall statistic was 0.84. A higher KMO value (0.60 or more) indicates that there is some underlying structure (i.e., common factor) in the data. A factor analysis estimates the degree to which the variability of observed variables is due to common factors. The proportion of variance of each variable that can be explained by common factors is called communality. The communalities derived from our data were less than 1, indicating that the observed variables were not highly correlated with each other. To determine the number of factors, we employed the Kaiser criterion, the 50% variance explained criterion, and a criterion based on parallel analysis. Parallel analysis, in which a factor is retained if its eigenvalue is larger than the 95th percentile of the distribution of eigenvalues derived from random data (Humphreys and Ilgen 1969), is one of the most highly recommended methods.

Confirmatory factor analysis is similar to exploratory factor analysis. However, the former allows researchers to determine factor models a priori on the basis of theoretical considerations, whereas the latter obtains the underlying factor model a posteriori. Using confirmatory factor analysis, we compared fit indices of models with 1, 2, 3, and 4 factors. A likelihood ratio test (e.g., χ^2 test) is frequently used when the fit of models is evaluated. However, the χ^2 statistic measures the badness of fit of an evaluated model compared with a saturated model:

The nonsignificant χ^2 statistic means a good fit. Previous studies have proposed several indices with which to evaluate the fit of models (Hu and Bentler 1998). The present study selected different types of fit indices: the Akaike information criterion (AIC), the standardized root mean-squared residual (SRMR), the Bollen incremental fit index (IFI), and the Bentler comparative fit index (CFI). The AIC and SRMR are absolute fit indices, whereas the IFI and CFI are relative fit and noncentrality-based indices, respectively. The AIC measures the complexity of the evaluated model in terms of the degree of freedom and penalizes more complex models. The SRMR, in the same way as the χ^2 statistic, assesses the difference between predicted and observed covariances but is insensitive to small sample sizes ($N < 150$). The IFI and CFI quantify the extent to which the evaluated model is better than a baseline model, in which correlations between observed variables are set at zero. Finally, we performed the χ^2 difference test to directly compare the fit between the models. In this test, the χ^2 statistic for the evaluated model is subtracted from the χ^2 statistic for a nested model with a larger degree of freedom.

For a genotype group comparison, we averaged the number of perceptual switches for the 2 conditions in each task and obtained the results of both a task-based and a condition-based comparison. As for the number of perceptual switches, we performed a one-way analysis of variance (ANOVA) and used Sidak multiple comparisons as a post hoc test (alpha level = 0.05). These results were essentially the same. For ease of interpretation, the task-based comparison is reported below.

Results

A histogram of the percept durations for each task fits a gamma distribution, which is consistent with previous findings on binocular rivalry (Levelt 1965; Fox and Herrmann 1967; Blake et al. 1971; Lehky 1995; Logothetis et al. 1996). It is known that perceptual predominance depends on Δf between H and L tones of a triplet in auditory streaming and on Δd between 2 gratings in visual plaids. The results showed that the total duration of grouped percepts was longer at $\Delta f = 2$ semitones than at $\Delta f = 6$ semitones; at $\Delta d = 100^{\circ}$ than at $\Delta d = 150^{\circ}$: $t_{91} = 9.65$ and 14.21, $P < 0.001$ (Table 1). However, the number of perceptual switches varied with the conditions.

We performed a correlation analysis on the number of perceptual switches. All the correlation coefficients between the conditions were positive (range 0.11–0.75) (Table 2). This is consistent with a previous finding, which revealed that there is a significant correlation between the number of switches for different modality tasks: the auditory streaming and visual plaid tasks ($r = 0.40$) (Pressnitzer and Hupé 2006). In addition, the present study showed that the correlations between the 2 conditions in each task were moderately high ($r > 0.43$). The within-task correlations were greater for the verbal transformation and reversible figure tasks than for the auditory streaming and visual plaid tasks: 0.75 and 0.69 (90% confidence

Table 1

Descriptive statistics for the auditory and visual tasks ($N = 92$)

Task	Condition	Number of switches	Grouped percept durations (s)
Auditory streaming	2 semitones	50.8 (3.9)	233.6 (6.2)
	6 semitones	52.6 (5.0)	158.6 (6.5)
Verbal transformations	Banana	68.4 (4.5)	N/A
	Token	63.6 (3.9)	N/A
Visual plaids	100 $^{\circ}$	63.3 (3.7)	311.6 (6.0)
	150 $^{\circ}$	88.8 (3.8)	211.1 (5.2)
Reversible figures	Necker cube	64.1 (3.2)	N/A
	Rubin vase	46.4 (3.4)	N/A

Note: The mean and standard error of mean (in parentheses) of five 90-s trials are shown. N/A, not applicable.

Table 2
Correlation matrix of the number of perceptual switches

Condition	1	2	3	4	5	6	7	8
1. AS2	—	0.45**	0.40**	0.65**	0.58**	0.38**	0.28**	0.34**
2. AS6		—	0.32**	0.50**	0.30**	0.32**	0.26**	0.35**
3. Banana			—	0.75**	0.49**	0.39**	0.34**	0.27**
4. Tokei				—	0.22	0.47**	0.33**	0.11
5. VP100					—	0.43**	0.31**	0.24*
6. VP150						—	0.37**	0.23*
7. Necker							—	0.69**
8. Rubin								—

Note: For the abbreviations, see Figure 2.
** $P < 0.01$, * $P < 0.05$.

interval 0.66–0.82 and 0.59–0.77); 0.45 and 0.43 (0.30–0.58 and 0.28–0.56). This discrepancy in the correlation coefficients of the tasks suggests that the temporal variability of the perceptual switches depends on the characteristics of the perceptual processes rather than on the modality of the sensory inputs, although the number of switches for each individual may change with the stimulus parameters.

We performed an exploratory factor analysis to identify the relationship between the 8 conditions. The results demonstrated that 3 factors, whose eigenvalues were larger than 1, were extracted from the conditions (Fig. 2). The first factor, with an eigenvalue of 3.25 before the rotation, was heavily loaded on the reversible figure tasks (factor loadings 0.87 and 0.66), whereas the second factor, with an eigenvalue of 1.14, was loaded on the visual plaid tasks (0.81 and 0.61). Thus, it can be interpreted that the first and second factors are the “shape” and “motion” factors, respectively. The third factor had an eigenvalue of 1.07 and was loaded on the verbal transformation and auditory streaming tasks (0.32–0.71), suggesting that this is the “auditory” factor. The 3 factors accounted for a 63.6% variance in the number of perceptual switches in all the data. Also, the parallel analysis results indicated that a 3-factor solution fitted the data.

Using a confirmatory factor analysis, we attempted to specify that the 3 factors, namely the “auditory,” “shape,” and “motion” factors, were separable at the perceptual task level but correlated with each other (Fig. 3C). For this purpose, we constructed other possible models in addition to the 3-factor model and compared the models’ fit indices. A 1-factor model assumed that the “general” factor accounted for variances of the number of switches in all the tasks (Fig. 3A). A 2-factor model postulated that the “auditory” and “visual” factors explained variances of the number of switches in the auditory and visual tasks, respectively (Fig. 3B). In a 4-factor model, the “nonspeech,” “speech,” “motion,” and “shape” factors corresponded to each task (Fig. 3D). The AIC, SRMR, IFI, and CFI values were generally better for the 3-factor model than for the other models (Table 3). The χ^2 difference test showed a significant result in the contrast of the 1- and 3-factor models, $\chi^2_{(3)} = 25.11$, $P < 0.01$; and in the contrast of the 2- and 3-factor models, $\chi^2_{(2)} = 12.39$, $P < 0.01$. This indicates that the entry of additional factor(s) is statistically meaningful. However, the χ^2 difference between the 3- and 4-factor models did not reach a significant level, $\chi^2_{(3)} = 2.07$, n.s. Thus, the 3-factor model is more parsimonious than the 4-factor model. Taken collectively, the results of exploratory and confirmatory factor analyses support the view that the 3-factor solution provides

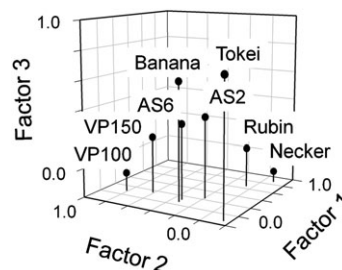


Figure 2. Results of exploratory factor analysis. The number of perceptual switches for each individual ($N = 92$) is used as an observed variable. The factors were extracted by using the maximum likelihood method and then subjected to an oblique Promax rotation, in which it is postulated that the factors behind the observed variables are not completely independent of one another. The circles are plotted on the basis of factor loadings for each of the 8 conditions. Factors 1, 2, and 3 can be considered the “shape,” “motion,” and “auditory” factors, respectively. AS2 and AS6, auditory streaming at $\Delta f = 2$ and 6 semitones; VP100 and VP150, visual plaids at $\Delta\theta = 100^\circ$ and 150° .

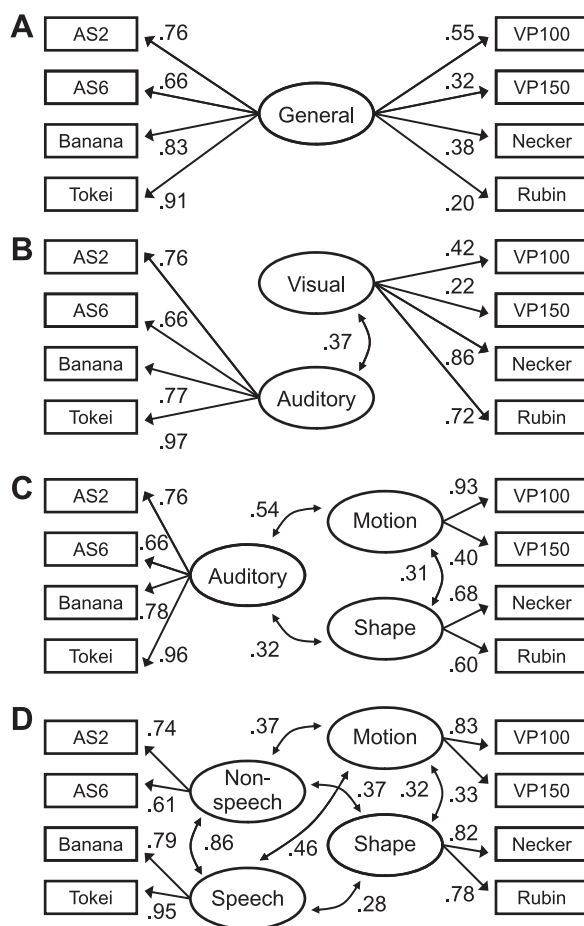


Figure 3. Results of confirmatory factor analysis. (A–D) One-, 2-, 3-, and 4-factor models. The 3-factor model is endorsed in terms of fit indices (see Table 3). The squares and ellipses represent observed and latent variables, respectively. The single-headed arrows have standardized regression coefficients that are equivalent to factor loadings, whereas the double-headed arrows indicate correlations between the latent variables. All of the coefficients are significant ($P < 0.05$).

a better fit to the data derived from auditory streaming, verbal transformations, visual plaids, and reversible figures.

For each perceptual task, we compared the number of perceptual switches among genotype groups to examine which neurotransmitter functions were associated with the

Table 3

Fit indices for the models in confirmatory factor analysis

Model	χ^2	df	<i>P</i>	χ^2/df	AIC	SRMR	IFI	CFI
One factor	58.70	20	0.001	2.94	106.70	0.125	0.74	0.72
Two factors	45.98	19	0.001	2.42	95.98	0.148	0.82	0.80
Three factors	33.59	17	0.009	1.98	87.59	0.080	0.90	0.88
Four factors	31.52	14	0.005	2.25	91.52	0.094	0.89	0.87

Note: Nonsignificant χ^2 statistics and $\chi^2/df < 2$ represent a good fit to the data. A lower AIC value indicates a better fit. A lower SRMR value (0.080 or less) and higher IFI and CFI values (0.90 or more) both indicate a better fit. df, degrees of freedom.

factors. An ANOVA revealed that the number of perceptual switches in the auditory streaming task was greater for the Met/Met group (mean \pm standard error of mean, 77.7 ± 14.0) than for the Val/Met group (46.7 ± 4.2) and Val/Val group (44.2 ± 5.1); $F_{2,89} = 5.72$, $\eta^2 = 0.013$, $P < 0.01$ (Fig. 4A). In the verbal transformation task, the number of perceptual switches was greater for the Met/Met group (90.2 ± 14.8) than for the Val/Met group (60.5 ± 5.7) and Val/Val group (66.2 ± 5.9); $F_{2,89} = 3.23$, $\eta^2 = 0.007$, $P < 0.05$ (Fig. 4A). In the reversible figure task, there was a marginal difference between the A/A group (62.5 ± 7.1) and G/G group (44.4 ± 4.7); $F_{2,88} = 2.36$, $\eta^2 = 0.004$, $P < 0.10$ (Fig. 4B). These results suggest that the number of perceptual switches in auditory streaming and verbal transformations is affected by the dopamine system, whereas that in reversible figures is influenced by the serotonin system. Taking the results of the factor analyses into account, we can expect the auditory and visual factors to reflect the neural processes of the dopamine and serotonin functions, respectively.

Discussion

The present results demonstrate that differences in neurotransmitter function lead to a wide range of individual variations in bistable perception. In this study, we collect switching rate data across different task types and specify the genotype/molecular causes of such variations. The number of perceptual switches is positively correlated between different forms of bistable stimuli. Furthermore, the results of factor analyses show that the 3 factors, namely the auditory, shape, and motion factors, are correlated with each other but distinguishable. The numbers of perceptual switches in auditory streaming and verbal transformations differ for different COMT genotype groups, whereas that in reversible figures differs for different HTR2A genotype groups. Our results indicate that the dopamine and serotonin systems are closely linked with auditory and shape factors. Below, we discuss the neural processes underlying the separability and commonality of bistable perception.

The results of factor analyses of the number of perceptual switches revealed that the 3 factors were extracted from the 8 conditions and were separable in terms of perceptual tasks (Fig. 2 and Table 3). This is consistent with previous findings of human neuroimaging studies indicating that auditory and visual bistability involves processes in different brain areas. Auditory streaming and verbal transformations produce activity in the auditory-related areas (Gutschalk et al. 2005; Kondo and Kashino 2009) and frontal areas (Kondo and Kashino 2007). Reversible figures and visual plaids lead to activity in the ventral occipital area (Kleinschmidt et al. 1998) and motion sensitive area hMT+ (Huk and Heeger 2001; Castelo-Branco et al. 2002). Thus, it appears that the separable factors correspond to sensory

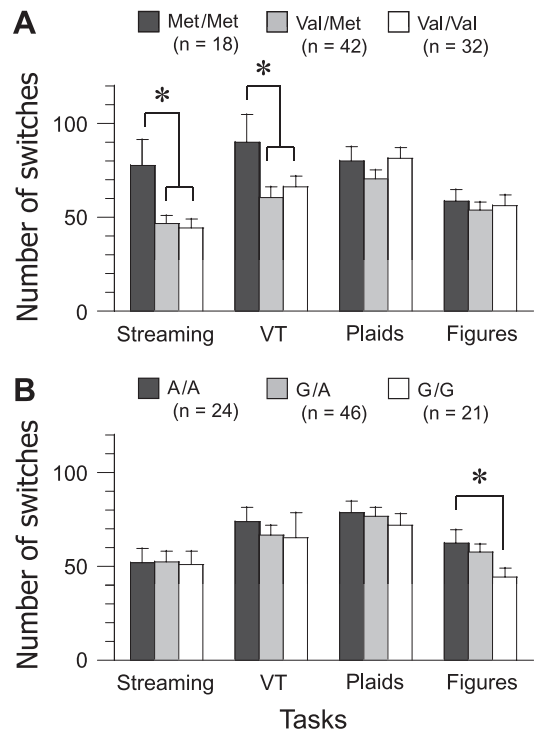


Figure 4. Number of perceptual switches for genotype groups. (A) COMT genotype group comparison. The number of perceptual switches was greater for the Met/Met group than for the Val/Met and Val/Val groups in the auditory streaming and verbal transformation tasks. Error bars represent the standard error of mean. (B) HTR2A genotype group comparison. The number of perceptual switches was greater for the A/A group than for the G/G group in the reversible figure task. * $P < 0.05$.

modules in the brain. In addition, we found that the COMT and HTR2A genetic effects on the number of perceptual switches were different for the auditory and visual tasks. Let us consider here why the dopamine and serotonin systems were involved differently in auditory and visual bistability in this study.

The number of perceptual switches in the auditory tasks was different for different COMT genotype groups, but this was not the case with the visual tasks (Fig. 4A). Dopamine is released from the brainstem nuclei and projected to the frontal and cingulate areas via the limbic areas (Cooper et al. 2003). The level of COMT expression is higher in the frontal and temporal cortices containing the auditory areas than in the occipital cortex containing the visual areas (Hong et al. 1998). A recent EEG study argued that the COMT genotype is closely linked with the sensory gating of auditory stimuli (Majic et al. 2011). Thus, the formation and selection of auditory percepts are probably more sensitive to dopamine signaling than the formation and selection of visual percepts. In addition, the enzymatic degradation of COMT containing Met is one-fourth as active as that containing Val (Lotta et al. 1995). Thus, it is likely that a lower metabolism and a higher concentration of dopamine in synapses enhance the signal-to-noise ratio of neurons (Weinberger et al. 2001), resulting in the following order for the number of perceptual switches: the Met/Met > Val/Met > Val/Val groups. In this study, auditory streaming and verbal transformations were extracted as the same auditory factor, although the stimuli of the 2 tasks were very different. The reason may be that auditory stimuli have a temporal structure but visual stimuli usually do not.

Temporal segmentation (i.e., distinctive rhythm patterns), as well as spectral grouping (i.e., 1 or 2 streams), is essential for resolving ambiguity in hearing. Thus, we believe that temporal segmentation results in the functional binding of the 2 auditory tasks.

The number of perceptual switches in reversible figures was different in different HTR2A genotype groups, but this was not the case in visual plaids (Fig. 4B). There is the possibility that the asymmetry of the results for the 2 tasks derives from the attributes of the stimuli. The visual system has separate pathways for processing shape and motion information even in the lateral geniculate nucleus (e.g., Friston 2005). Thus, the serotonin system may have greater effects on the perceptual formation of static objects than on that of moving objects. A few important studies have shown the close relationship between the serotonin system and visual bistability. The switching rate in binocular rivalry is decreased by the administration of psilocybin (Carter et al. 2005). This effect is not induced by the administration of psilocybin with ketanserin (HTR2A antagonist) (Carter et al. 2007) but is induced by the administration of tandospirone (HTR1A agonist) (Nagamine et al. 2008). The previous findings suggest that the enhancement of HTR1A, rather than of HTR2A, suppresses the binocular rivalry rate. Thus, future studies should try to clarify the effects of serotonin neurotransmission on visual bistability.

An early study revealed genetic effects on illusory movement, which occurs when observers view shaded stripes peripherally (Fraser and Wilcox 1979). A recent twin study revealed that additive genetic factors account for an approximately 50% variance in the binocular rivalry rate (Miller et al. 2010). Thus, we can expect bistable perception to be more or less influenced by genetic factors. However, although the interest of researchers has concentrated on visual bistability, particularly binocular rivalry, we have demonstrated the different involvement of the dopamine and serotonin systems in relation to auditory and visual bistability. Therefore, the present finding warrants a further elaboration of the bistable perception mechanisms and contributes to an understanding of the link between neurotransmitter function and human perception.

Thus far, we have discussed the neural processes underlying auditory and visual bistability. The genotype group comparison results seem to support the idea that perceptual switching mechanisms are distributed at the cortical level (Pressnitzer and Hupé 2006; Hupé et al. 2008). However, a possible interpretation is that our results were induced by genetic effects on developmental processes. The middle frontal and anterior temporal cortices are sensitive to genetic influences on brain structures (Toga and Thompson 2005). The gray matter volumes of the inferior frontal and superior temporal cortices are correlated with differences in intellectual function as quantified by *g*, which is also highly heritable (Thompson et al. 2001). Thus, variations in cortical structures and functions during development may lead to genotype group differences in the switching rate between auditory and visual bistability. However, it should be noted that we found significant positive correlations between the numbers of perceptual switches even under different modality conditions (Table 2). This suggests that some common factor mediates different forms of bistable perception. Thus, it is difficult to explain this commonality solely in terms of genetic effects on cortical structures and functions.

There is the possibility that the commonality of auditory and visual bistability is derived from the subcortical function because dopamine and serotonin are released from brainstem nuclei, such as the ventral tegmental area and dorsal raphe. This idea is consistent with a central oscillator hypothesis in which biological rhythms generated by the brainstem modulate the temporal dynamics of perceptual switches (Pettigrew and Miller 1998; Carter and Pettigrew 2003; Sheppard and Pettigrew 2006). Intriguingly, this hypothesis postulates that multiple oscillators drive interhemispheric synchronous neural processing. In light of our results, we can expect separate oscillators to produce rhythms for the perceptual switching of auditory and visual bistability and to be connected with each other at the brainstem level. Several researchers have also pointed out that the brainstem nuclei play an important role in switching timing regardless of auditory and visual modalities. A recent study has revealed that pupil size increases before perceptual switches in different ambiguous stimuli, suggesting that the autonomic nervous system affects bistable perception because pupil dilation is regulated by noradrenaline from the locus coeruleus (Einhäuser et al. 2008). That study argues that pupil dilation is an important predictor of percept duration, but the major component of pupil dilation may be induced by the button press used to indicate perceptual switches (Hupé et al. 2009). Thus, showing that the exact role of the brainstem nuclei as regards the commonality requires further study.

We can also consider top-down modulation as a candidate for the commonality that influences the individual variability of perceptual switches. It has been pointed out that endogenous attention plays a general role in switching from one percept to another and in maintaining perceptual stability (Leopold and Logothetis 1999). Attentional effects on bistable perception have been investigated in auditory science (Carlyon et al. 2001; Cusack 2005; Snyder et al. 2006) and visual science (Suzuki and Peterson 2000; Meng and Tong 2004; van Ee et al. 2005). In addition, brain stimulation of attention-related areas changes the predominance time of one possible image during binocular rivalry and reversible figures (Miller et al. 2000; Ngo et al. 2007). However, psychophysical studies have demonstrated that attentional effects on the number of perceptual switches are weaker for binocular rivalry than for the Necker cube (Meng and Tong 2004) and that the degree of volitional control that biases perceptual dominance is different for auditory streaming and visual plaids in each individual (Pressnitzer and Hupé 2006). Thus, although endogenous attention seems to influence the temporal dynamics of bistable perception, its effects may vary with different stimuli and modalities across individuals.

The present study includes some limitations. We selected 2 candidate genes, but the HTR2A genotype was only sensitive to individual variations in the number of perceptual switches in reversible figures. We also genotyped a variant of the serotonin transporter linked polymorphic region (5-HTTLPR) (see Koizumi et al. 2010). The 5-HTTLPR is a variable repeat sequence (i.e., short and long alleles) in the promoter region of the gene and its variations affect emotional traits (Hariri and Holmes 2006). We found a trend showing that the number of perceptual switches was greater for short-allele carriers regardless of the bistable perception phenomena used in this study, although we did not report the results due to the small sample size of long/long individuals. This may indicate that there is a single gene influencing bistable perception in

general. Thus, it is important to look for other genes related to individual differences in bistable perception as well as replicate the candidate gene analyses.

In contrast to previous studies investigating the characteristics of a specific bistable stimulus, the present study obtained large-scale sample data using different forms of auditory and visual bistable perception. Compared with the effects of genetic variations on higher-level cognition, little attention has been paid to the effects of genetic variations on lower-level perception. We have demonstrated that separable but correlated factors can be extracted from the perceptual switch data and closely linked with COMT and HTR2A genetic variations. The results suggest that individual differences in perceptual switching are induced by the functions of the dopamine and serotonin systems and that the formation and selection of percepts involve neural processes in cortical and subcortical areas.

Notes

We thank Jean-Michel Hupé and 2 anonymous reviewers for their thoughtful comments on an earlier version of this manuscript. *Conflict of Interest*: None declared.

References

- Adelson EH, Movshon JA. 1982. Phenomenal coherence of moving visual patterns. *Nature*. 300:523–525.
- Aghajanian GK, Marek GJ. 1997. Serotonin induces excitatory postsynaptic potentials in apical dendrites of neocortical pyramidal cells. *Neuropharmacology*. 36:589–599.
- Blake R, Fox R, McIntyre C. 1971. Stochastic properties of stabilized-image binocular rivalry alternations. *J Exp Psychol*. 88:327–332.
- Bregman AS. 1990. Auditory scene analysis: the perceptual organization of sound. Cambridge (MA): MIT Press.
- Carlyon RP, Cusack R, Foxtton JM, Robertson IH. 2001. Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform*. 27:115–127.
- Carter OL, Hasler F, Pettigrew JD, Wallis GM, Liu GB, Vollenweider FX. 2007. Psilocybin links binocular rivalry switch rate to attention and subjective arousal levels in humans. *Psychopharmacology (Berl)*. 195:415–424.
- Carter OL, Pettigrew JD. 2003. A common oscillator for perceptual rivalries? *Perception*. 32:295–305.
- Carter OL, Pettigrew JD, Hasler F, Wallis GM, Liu GB, Hell D, Vollenweider FX. 2005. Modulating the rate and rhythmicity of perceptual rivalry alternations with the mixed 5-HT_{2A} and 5-HT_{1A} agonist psilocybin. *Neuropsychopharmacology*. 6:1154–1162.
- Castelo-Branco M, Formisano E, Backes W, Zanella F, Neuenschwander S, Singer W, Goebel R. 2002. Activity patterns in human motion-sensitive areas depend on the interpretation of global motion. *Proc Natl Acad Sci U S A*. 99:13914–13919.
- Cooper JR, Bloom FE, Roth RH. 2003. The biochemical basis of neuropharmacology. New York: Oxford University Press.
- Cusack R. 2005. The intraparietal sulcus and perceptual organization. *J Cogn Neurosci*. 17:641–651.
- Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR. 2001. Effect of COMT Val^{108/158} Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci U S A*. 98:6917–6922.
- Einhäuser W, Stout J, Koch C, Carter O. 2008. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proc Natl Acad Sci U S A*. 105:1704–1709.
- Fox R, Herrmann J. 1967. Stochastic properties of binocular rivalry alternations. *Percept Psychophys*. 2:432–436.
- Fraser A, Wilcox KJ. 1979. Perception of illusory movement. *Nature*. 281:565–566.
- Friston K. 2005. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*. 360:815–836.
- Goldberg TE, Egan MF, Gscheidle T, Coppola R, Weickert T, Kolachana BS, Goldman D, Weinberger DR. 2003. Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype schizophrenia. *Arch Gen Psychiatry*. 60:889–896.
- Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. 2005. Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci*. 25:5382–5388.
- Hariri AR, Holmes A. 2006. Genetics of emotional regulation: the role of the serotonin transporter in neural function. *Trends Cogn Sci*. 10:182–191.
- Haynes JD, Deichmann R, Rees G. 2005. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature*. 438:496–499.
- Hong J, Shu-Leong H, Tao X, Lap-Ping Y. 1998. Distribution of catechol-O-methyltransferase expression in human central nervous system. *Neuroreport*. 9:2861–2864.
- Hu L, Bentler PM. 1998. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Methods*. 3:424–453.
- Huk AC, Heeger DJ. 2001. Pattern-motion responses in human visual cortex. *Nat Neurosci*. 5:72–75.
- Humphreys LG, Ilgen DL. 1969. Note on a criterion for the number of common factors. *Educ Psychol Meas*. 29:571–578.
- Hupé JM, Joffo LM, Pressnitzer D. 2008. Bistability for audiovisual stimuli: perceptual decision is modality specific. *J Vis*. 8:1–15.
- Hupé JM, Lamirel C, Lorenceau J. 2009. Pupil dynamics during bistable motion perception. *J Vis*. 9:1–19.
- Hupé JM, Rubin N. 2003. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res*. 43:531–548.
- Jakab RL, Goldman-Rakic PS. 1998. 5-Hydroxytryptamine_{2A} serotonin receptors in the primate cerebral cortex: possible site of action of hallucinogenic and antipsychotic drugs in pyramidal cell apical dendrites. *Proc Natl Acad Sci U S A*. 95:735–740.
- Kleinschmidt A, Büchel C, Zeki S, Frackowiak RSJ. 1998. Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc R Soc Lond B Biol Sci*. 265:2427–2433.
- Koizumi A, Kitagawa N, Kitamura MS, Kondo HM, Sato T, Kashino M. 2010. Serotonin transporter gene and inhibition of conflicting emotional information. *Neuroreport*. 21:422–426.
- Kondo HM, Kashino M. 2007. Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage*. 36:123–130.
- Kondo HM, Kashino M. 2009. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J Neurosci*. 29:12695–12701.
- Lehky SR. 1995. Binocular rivalry is not chaotic. *Proc R Soc Lond B Biol Sci*. 259:71–76.
- Leopold DA, Logothetis NK. 1999. Multistable phenomena: changing views in perception. *Trends Cogn Sci*. 3:254–264.
- Levelt WJM. 1965. On binocular rivalry. Soesterberg (the Netherlands): Institute for Perception RVO-TNO.
- Logothetis NK, Leopold DA, Sheinberg DL. 1996. What is rivaling during binocular rivalry? *Nature*. 380:621–624.
- Long GM, Toppino TC. 2004. Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychol Bull*. 130:748–768.
- Lotta T, Vidgren J, Tilgmann C, Ulmanen I, Melén K, Julkunen I, Taskinen J. 1995. Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry (Mosc)*. 34:4202–4210.
- Majic T, Rentzsch J, Gudowski Y, Ehrlich S, Juckel G, Sander T, Lang UE, Winterer G, Gallinat J. 2011. COMT Val^{108/158} Met genotype modulates human sensory gating. *Neuroimage*. 55:818–824.
- Meng M, Tong M. 2004. Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J Vis*. 4:539–551.
- Miller SM, Hansell NK, Ngo TT, Liu GB, Pettigrew JD, Martin NG, Wright MJ. 2010. Genetic contribution to individual variation in binocular rivalry rate. *Proc Natl Acad Sci U S A*. 107:2664–2668.
- Miller SM, Liu GB, Ngo TT, Hooper G, Riek S, Carson RG, Pettigrew JD. 2000. Interhemispheric switching mediates perceptual rivalry. *Curr Biol*. 10:383–392.

- Nagamine M, Yoshino A, Miyazaki M, Takahashi Y, Nomura S. 2008. Effects of selective 5-HT_{1A} agonist tandospirone on the rate and rhythmicity of binocular rivalry. *Psychopharmacology (Berl)*. 198:279-286.
- Ngo TT, Liu GB, Tilley AJ, Pettigrew JD, Miller SM. 2007. Caloric vestibular stimulation reveals discrete neural mechanisms for coherence rivalry and eye rivalry: a meta-rivalry model. *Vision Res*. 47:2685-2699.
- Pettigrew JD, Miller SM. 1998. A 'sticky' interhemispheric switch in bipolar disorder? *Proc R Soc Lond B Biol Sci*. 265:2141-2148.
- Pressnitzer D, Hupé JM. 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol*. 16:1351-1357.
- Schadwinkel S, Gutschalk A. 2011. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J Neurophysiol*. 105:1977-1983.
- Shannon RW, Patrick CJ, Jiang Y, Bernat E, He S. 2011. Genes contribute to the switching dynamics of bistable perception. *J Vis*. 11:1-7.
- Sheppard BM, Pettigrew JD. 2006. Plaid motion rivalry: correlates with binocular rivalry and positive mood state. *Perception*. 35:157-169.
- Snyder JS, Alain C, Picton TW. 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci*. 18:1-13.
- Suzuki S, Peterson MA. 2000. Multiplicative effects of intention on the perception of bistable apparent motion. *Psychol Sci*. 11:202-209.
- Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen VP, Huttunen M, Lönqvist J, Standertskjöld-Nordenstam CG, Kaprio J, Khaledy M, et al. 2001. Genetic influences on brain structure. *Nat Neurosci*. 12:1253-1258.
- Toga AW, Thompson PM. 2005. Genetics of brain structure and intelligence. *Annu Rev Neurosci*. 28:1-23.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N. 1998. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*. 21:753-759.
- van Ee R, van Dam LCI, Brouwer GJ. 2005. Voluntary control and the dynamics of perceptual bi-stability. *Vision Res*. 45:41-55.
- van Noorden LPAS. 1975. Temporal coherence in the perception of tone sequences. [PhD thesis] [Eindhoven (the Netherlands)]: Eindhoven University of Technology.
- Wallach H. 1935. Über visuell wahrgenommene Bewegungsrichtung. *Psychol Forsch*. 20:325-380[Wuerger S, Shapley R, Rubin N. (Trans.). 1996. "On the visually perceived direction of motion" by Hans Wallach: 60 years later. *Perception*. 25:1317-1367].
- Warren RM. 1961. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *Br J Psychol*. 52:249-258.
- Warren RM, Gregory RL. 1958. An auditory analogue of the visual reversible figure. *Am J Psychol*. 71:612-613.
- Weinberger DR, Egan MF, Bertolino A, Callicott JH, Mattay VS, Lipska BK, Berman KF, Goldberg TE. 2001. Prefrontal neurons and the genetics of schizophrenia. *Biol Psychiatry*. 50:825-844.
- Wunderlich K, Schneider KA, Kastner S. 2005. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci*. 8:1595-1602.