Letters to the Editor

Reply to: Evaluating the role of the cerebellum in temporal processing: beware of the null hypothesis

Deborah L. Harrington[1] Roland R. Lee[2] Lara A. Boyd[3] Steven Z. Rapcsak[4] and Robert T. Knight[5]

[1]New Mexico Veterans Affairs Health Care System and Department of Neurology, University of New Mexico Health Sciences Center, Albuquerque, NM, [2]Department of Radiology, University of California, San Diego, CA, [3]Departments of Neurology and Rehabilitation Sciences, University of Kansas Medical Center, Kansas City, KS, [4]Southern Arizona Veterans Affairs Health Care System and Department of Neurology, University of Arizona, Tucson, AZ and [5]Department of Psychology and Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA, USA

Correspondence to: Deborah L. Harrington, Psychology (116B), Veterans Affairs Medical Center, 1501 San Pedro SE, Albuquerque, NM 87108, USA. E-mail: dharring@unm.edu
DOI: 10.1093/brain/awh227

The neural systems that regulate temporal aspects of behaviours within the range of several hundred milliseconds, several seconds and even longer periods remain debated. Parametric studies of timing across these interval ranges, which seem to have psychological significance, have not yet been carried out by any single study using neuroscience methods. Most research has studied patients with neurological damage on a single timing task, typically testing only one interval. This body of work led to the influential neuropsychological model of Ivry and colleagues, in which the cerebellum is viewed as a central timekeeping mechanism that computes time for intervals in the range of several hundred milliseconds (Ivry *et al*., 1988; Ivry and Keele, 1989; Ivry, 1996). The strongest support for this proposal comes from studies demonstrating that patients with cerebellar damage are impaired on tests in which they must explicitly estimate or reproduce the duration of an interval lasting several hundred milliseconds. As we pointed out in our paper, evidence for the cerebellar timing model has been limited in primarily two ways. First, most studies have included patients with cerebellar atrophy. The use of these patients to draw direct inferences about the role of the cerebellum in any behaviour is limited because degenerative cerebellar atrophy is rarely restricted to the cerebellum and may damage other brain regions. Indeed, cerebellar atrophy patients show marked deficits in temporal processing relative to patients with cerebellar damage due to stroke (Casini and Ivry, 1999). This problem cannot be overstated, given that damage to various areas of the cerebral cortex also produces deficits in timing intervals in the range of hundreds of milliseconds and seconds (Nichelli *et al*., 1995; Rubia *et al*., 1997; Harrington *et al*., 1998; Mangels *et al*., 1998; Casini and Ivry, 1999). These data suggest at least two possibilities. First, timing may be centralized, but the role of various brain regions is difficult to distinguish because experimental methods do not adequately differentiate deficits in timekeeping mechanisms from other processes that support timing (e.g. memory, attention). Alternatively, timing may be a distributed process supported by more than one brain region. A second limitation is that most studies have not provided evidence that cerebellar damage produces timing deficits across more than one interval and/or more than one test of timing. This is important because prevailing theories maintain that a common timekeeping mechanism supports timing in different tasks and different intervals within a similar psychological range. For these reasons, we set out to provide a stronger test of the cerebellar timing hypothesis by studying a large sample of patients with chronic cerebellar damage due to stroke and examining performance on two different timing tasks, each of which contained two different standard interval conditions. We hypothesized that patients should show deficits on all four conditions if the cerebellum regulates a timekeeping mechanism.

Ivry and Spencer take issue with our interpretation of the pattern of temporal processing performance in cerebellar patients on several grounds. First, they maintain there is broad support for the cerebellar timing model because the cerebellum also plays a role in other tasks that appear to involve timing (e.g. throwing, speech, speech perception, vestibulo-ocular adaptation). We believe that these data constitute only indirect evidence because it is not clear that the cerebellum's role in these tasks is one of timekeeping. Using this criterion, an equally strong case could be made for the role of other brain regions in timing. Although we agree that evidence for a brain region's role in regulating timing should come from a wide range of tasks, we would add that there needs to be evidence that explicit timing is engaged by those tasks.

Ivry and Spencer's main critique is that it is premature to accept the null hypothesis on the basis of a 'marginally significant result' that is not in accord with previous findings. Here they refer to our statistically non-significant difference (P = 0.07) between controls and patients with superior cerebellar

lesions on the time perception task. The term 'marginally significant' is problematic on a number of accounts; for example, it erroneously implies that a small value of P means that there is an effect of a substantial magnitude to warrant theoretical significance, or makes a statement about the ability to replicate the results (Nickerson, 2000). The implicit supposition is that we should abandon the conventional criterion for establishing sufficient evidence to support a claim, where sufficient is defined as $P < 0.05$. The criterion is not that P should be small, but that it should be less than 0.05 (Frick, 1996). It is important to emphasize that P is not the probability that the null hypothesis is correct, and its complement $1 - P$ is not the probability that the alternative hypothesis is correct. Rather, P represents the probability of obtaining the observed results (or larger) given the assumption that the null hypothesis is true. Thus, if P is set at 0.05, it is treated the same as a test statistic with a value of 0.001, and $P = 0.07$ is treated the same as 0.70. Several studies have demonstrated that there is a good justification for the 0.05 criterion because it exhibits a 'cliff characteristic', wherein the reported confidence in a finding drops abruptly when P becomes larger than this value (Frick, 1996; Nickerson, 2000). Although it has been argued that investigators should be able to set their own rules for statistical significance testing, this practice has not prevailed because it introduces subjectivity into hypothesis testing (Chow, 1988) and increases the risk of basing hypothesis testing on the basis of personal belief, rather than on the strength of evidence (Frick, 1996). Of particular relevance to these issues is that we did not adjust our P level for multiple comparisons, as is the convention. No adjustment was made in the interest of granting the alternative hypothesis an 'edge', given the existing model. Thus, in our study the statistical criterion used for significance errs in the direction of mistakenly accepting the alternative hypothesis (type II error).

Having said this, we agree that a P value close to the 0.05 criterion should not be the only criterion by which a claim is made for failing to reject the null hypothesis, just as its complement $1 - P$ should not be the only criterion for accepting the alternative hypothesis. One type of information that augments null hypothesis statistical testing is individual subject data. Table 1 describes the performance of our cerebellar patients across the four timing conditions as a function of whether their performance fell outside the upper bound of the 95% confidence intervals in the control group or was within normal limits (wnl). This table shows that only three subjects (S9, S10, S21) exhibited impaired time perception and reproduction in both interval conditions of each task. Another three subjects showed impaired performance in three of the four conditions (S3, S11, S18). Even if we were to allow that this latter group was actually impaired on all conditions, the data indicate that a maximum of 28% of the 21 total patients showed a pattern of performance consistent with the cerebellar timing hypothesis. Only four of the 11 patients (36%) with lesions that extended into superior portions of the cerebellum were impaired in three or all four conditions. Thus, like our group analyses, the pattern of performance in individual subjects does not provide compelling support for the cerebellar timing hypothesis.

The inclination to accept the null hypothesis also depends on the goodness of the attempt to find an effect of cerebellar damage on temporal information processing. Here it is important to emphasize that our study was not a mere replication of Ivry and colleagues' work. Rather, we studied patients with chronic cerebellar strokes to control for confounding interpretations of data in studies using patients with cerebellar atrophy. We also studied performance across four conditions, because evidence for timing deficits should come from a range of tasks in which there is consensus that explicit timing mechanisms are involved. In addition, we tested three times as many chronic stroke patients as other studies to increase statistical power. Thus, our study provided a stronger and more direct test of the cerebellar timing hypothesis. Moreover, our findings were not simply null results, but instead demonstrated that superior cerebellar lesions disrupted time reproduction in the impaired limb, but not time perception.

Another issue pertains to the viability of the cerebellar timing model as it is presently construed. Ivry and Spencer maintain that it is reasonable to assume a 'dual-pattern of impairment' in patients with lesions to the superior aspect of the cerebellum, because the 'marginally significant' effect is 'indicative of deficit' when 'coupled with previous reports of elevated perceptual thresholds on similar time perception tasks in patients with cerebellar lesions'. We are not aware of any study of time perception that has separately examined performance in patients with chronic cerebellar lesions due to stroke using statistical verification procedures. Rather, Ivry and colleagues have described elevated perceptual thresholds in a group of seven patients with acute stroke or tumor resections, but normal time perception performance in a group of seven patients with chronic stroke or tumor resection (Ivry *et al*., 1988). Ivry and Spencer go on to say that, even if they accept the null hypothesis in our study, our time perception findings are still consistent with their model. In their model, the cerebellum contains duplicate timing circuitry in each hemisphere. After unilateral cerebellar damage, the opposite cerebellar hemisphere takes over the timing functions, so that patients with unilateral cerebellar damage should show normal time perception performance. It appears that their model can predict both normal and 'minimally' impaired time perception performance in patients with unilateral damage. This is a problem because the model is difficult to test. Their model also predicts that bilateral damage should disrupt time perception. Table 1 shows that patients with bilateral damage (S3, S5, S17, S19) were not more likely than patients with unilateral damage to show greater time-perception deficits, despite some bilateral damage to more superior portions of the cerebellum in all patients. Rather, five out of six patients who showed time-perception deficits in both standard interval conditions had unilateral damage (S6, S9, S10, S18, S21).

Of relevance to the above issue is that we usually ascribe functional significance to an area based on frank deficits in a particular behaviour, which often persist. For example,

Table 1 *Time perception and reproduction performance in patients with cerebellar lesions*

Cerebellar group	Subject	Time perception: difference threshold (ms)		Time reproduction: clock variability (ms)	
		300 ms	600 ms	300 ms	600 ms
Right inferior	S1	wnl	72	wnl	wnl
	S2	wnl	Wnl	wnl	wnl
	S3*	wnl	72	19	27
	S4	wnl	138	wnl	wnl
Right superior	S5*	wnl	Wnl	wnl	27
	S6	51	90	wnl	wnl
	S7	wnl	Wnl	wnl	wnl
	S8	wnl	Wnl	36	wnl
	S9	60	111	22	43
	S10	120	114	27	32
Left inferior	S11	48	Wnl	22	39
	S12	wnl	Wnl	wnl	wnl
	S13	wnl	72	wnl	27
	S14	wnl	Wnl	wnl	wnl
	S15	wnl	Wnl	wnl	wnl
	S16	wnl	Wnl	wnl	wnl
Left superior	S17*	wnl	Wnl	wnl	30
	S18	63	78	19	wnl
	S19*	48	72	wnl	wnl
	S20	wnl	84	wnl	27
	S21	78	78	17	43

*Time perception and reproduction performance is tabulated for subjects with cerebellar damage who showed impaired difference threshold or clock variability on one or more of the timing tasks (Harrington *et al*., 2004). Impairment was defined as performance greater than the upper bound of the 95% confidence interval in the control group. The upper bound for the time perception task was 46 and 68 ms for the 300 and 600 ms standard interval conditions, respectively. The upper bound clock variability was 15 and 26 ms for the 300 and 600 ms standard interval conditions, respectively. Performance that fell within normal limits is designated as wnl. Dark grey shading highlights patients who showed impaired performance in all four timing conditions. Light grey shading highlights patients who showed impaired performance in three timing conditions. Asterisks designate patients with bilateral cerebellar damage.

unilateral lesions to the prefrontal cortex from an ischaemic event can produce deficits in cognitive processes that are not strongly lateralized in humans, including working memory and executive functions (e.g. problem solving). While there may be some recovery or reorganization of function depending on the amount of tissue damage, these types of deficits commonly persist to some degree many years after a stroke; the intact hemisphere does not fully compensate for these deficits (Lezak, 1995). Likewise, time-perception deficits persisted in patients with right hemisphere damage who on the average were tested more than 4 years after stroke (Harrington *et al*., 1998). Ivry and Spencer take the opposite position, claiming that the cerebellum has the capacity to reorganize timing functions, so that the intact cerebellar hemisphere takes over. This is their explanation for why they find time-perception deficits in acute stroke patients, but not necessarily chronic stroke patients. Although we agree that both cerebellar hemispheres probably support the same functional mechanism, recovery of function can occur for many reasons, including the resolution of edema, recovery of function in the same region due to sparing of tissue, or reorganization of function within or between the hemispheres. As we pointed out in our paper, it is not necessary to assume that recovery is due to the intact hemisphere taking over. Moreover, as the model is construed it does not easily explain why the 'noisy temporal representation' from the damaged cerebellar hemisphere is readily combined with the normal representation in the intact hemisphere to produce normal or near-normal time perception, yet it is not utilized to reproduce intervals with the impaired hand. In the latter case, the intact hemisphere only helps out when both hands are working together. As we suggested in our paper, other explanations of these data are more parsimonious and do not require assumptions about the specific mechanism of recovery (Kelso *et al*., 1979; Breukelaar and Dalrymple-Alford, 1999).

The cerebellum's function in behaviour has long been debated because it plays a role in a wide range of tasks, many of which have no apparent explicit timing requirements. Our results in patients with chronic unilateral cerebellar stroke may not be due simply to recovery or reorganization of function, but also relate to the cerebellum's role in behaviour. Many models have been put forth to explain the cerebellum's role in sensorimotor and cognitive behaviours. The challenge is to develop models that can be tested straightforwardly and devise good experiments that can test between different models. In this regard, it is important to consider empirical findings in the broader human and animal literature, which show that damage to cortical and subcortical structures produces persistent deficits in explicit timing. While the reasons for this are not understood, it is clear that multiple brain systems participate in temporal information processing. Our data (Harrington

et al., 2004) also suggest that a better understanding of the cerebellum's role in timing or other behaviours may come from considering the functional significance of different regions within the cerebellum. Emerging neuroanatomical delineations of cortical-cerebellar input and output circuitry (Middleton and Strick, 2001; Dum and Strick, 2003) will undoubtedly help inform neuroscientists about their potential functional significance.

References

Breukelaar JW, Dalrymple-Alford JC. Effects of lesions to the cerebellar vermis and hemispheres on timing and counting in rats. Behav Neurosci 1999; 113: 78–90.

Casini L, Ivry R. Effects of divided attention on temporal processing in patients with lesions of the cerebellum or frontal lobe. Neuropsychology 1999; 13: 10–21.

Chow SL. Significance test or effect size? Psychol Bull 1988; 103: 105–10.

Dum RP, Strick PL. An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. J Neurophysiol 2003; 89: 634–9.

Frick RW. The appropriate use of null hypothesis testing. Psychol Methods 1996; 1: 379–90.

Harrington DL, Haaland KY, Knight RT. Cortical networks underlying mechanisms of time perception. J Neurosci 1998; 18: 1085–95.

Harrington DL, Lee RR, Boyd LA, Rapcsak SZ, Knight RT. Does the representation of time depend on the cerebellum? Effect of cerebellar stroke. Brain 2004; 127: 561–74.

Ivry R. The representation of temporal information in perception and motor control. Curr Opin Neurobiol 1996; 6: 851–7.

Ivry RB, Keele SW. Timing functions of the cerebellum. J Cogn Neurosci 1989; 1: 136–52.

Ivry RB, Keele SW, Diener HC. Dissociation of the lateral and medial cerebellum in movement timing and movement execution. Exp Brain Res 1988; 73: 167–80.

Kelso JA, Southard DL, Goodman D. On the coordination of two-handed movements. J Exp Psychol Hum Percept Perform 1979; 5: 229–38.

Lezak MD. Neuropsychological assessment. 3rd ed. New York: Oxford University Press; 1995.

Mangels JA, Ivry R, Shimizu N. Dissociable contributions of the prefrontal and neocerebellar cortex to time perception. Brain Res Cogn Brain Res 1998; 7: 15–39.

Middleton FA, Strick PL. Cerebellar projections to the prefrontal cortex of the primate. J Neurosci 2001; 21: 700–12.

Nichelli P, Clark K, Hollnagel C, Grafman J. Duration processing after frontal lobe lesions. Ann NY Acad Sci 1995; 769: 183–90.

Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. Psychol Methods 2000; 5: 241–301.

Rubia K, Schuri U, von Cramon DY, Poeppel E. Time estimation as a neuronal network property: a lesion study. Neuroreport 1997; 8: 1273–6.