A THEORY OF CEREBELLAR CORTEX

By DAVID MARR*

From Trinity College, Cambridge

SUMMARY

1. A detailed theory of cerebellar cortex is proposed whose consequence is that the cerebellum learns to perform motor skills. Two forms of input–output relation are described, both consistent with the cortical theory. One is suitable for learning movements (actions), and the other for learning to maintain posture and balance (maintenance reflexes).

2. It is known that the cells of the inferior olive and the cerebellar Purkinje cells have a special one-to-one relationship induced by the climbing fibre input. For learning actions, it is assumed that:

(*a*) each olivary cell responds to a cerebral instruction for an elemental movement. Any action has a defining representation in terms of elemental movements, and this representation has a neural expression as a sequence of firing patterns in the inferior olive; and

(*b*) in the correct state of the nervous system, a Purkinje cell can initiate the elemental movement to which its corresponding olivary cell responds.

3. Whenever an olivary cell fires, it sends an impulse (via the climbing fibre input) to its corresponding Purkinje cell. This Purkinje cell is also exposed (via the mossy fibre input) to information about the context in which its olivary cell fired; and it is shown how, during rehearsal of an action, each Purkinje cell can learn to recognize such contexts. Later, when the action has been learnt, occurrence of the context alone is enough to fire the Purkinje cell, which then causes the next elemental movement. The action thus progresses as it did during rehearsal.

4. It is shown that an interpretation of cerebellar cortex as a structure which allows each Purkinje cell to learn a number of contexts is consistent both with the distributions of the various types of cell, and with their known excitatory or inhibitory natures. It is demonstrated that the mossy fibre-granule cell arrangement provides the required pattern discrimination capability.

5. The following predictions are made.

(*a*) The synapses from parallel fibres to Purkinje cells are facilitated by the conjunction of presynaptic and climbing fibre (or post-synaptic) activity.

* Now at the Institute of Psychiatry, London, S.E. 5

(b) No other cerebellar synapses are modifiable.

(c) Golgi cells are driven by the greater of the inputs from their upper and lower dendritic fields.

6. For learning maintenance reflexes, 2(a) and 2(b) are replaced by

2′. Each olivary cell is stimulated by one or more receptors, all of whose activities are usually reduced by the results of stimulating the corresponding Purkinje cell.

7. It is shown that if (2′) is satisfied, the circuit receptor → olivary cell → Purkinje cell → effector may be regarded as a stabilizing reflex circuit which is activated by learned mossy fibre inputs. This type of reflex has been called a learned conditional reflex, and it is shown how such reflexes can solve problems of maintaining posture and balance.

8. 5(a), and either (2) or (2′) are essential to the theory: 5(b) and 5(c) are not absolutely essential, and parts of the theory could survive the disproof of either.

§0. INTRODUCTION

The cortex of the vertebrate cerebellum has a simple and extremely regular fine structure. This happy combination has made detailed experimental investigations possible, with the result that the arrangement and connexions of the cerebellar cells, together with the excitatory or inhibitory nature of the various synapses, are now clear (see Eccles, Ito & Szentagothai, 1967).

The structure of cerebellar cortex, though well understood, has as yet received no plausible interpretation. In the present paper, a theory of the cortex is proposed which explains what is known about it, and makes certain definite and testable predictions. The implication of the cortical theory is that the purpose of the cerebellum is to learn motor skills, so that when they have been learned a simple or incomplete message from the cerebrum will suffice to provoke their execution. Brindley (1964) suggested this was the function of the cerebellum.

The exposition is divided into various sections. In the first, an outline of the theory is presented: this is intended to provide a framework within which the reader may fit the details. The next five sections contain a cell by cell account of the cortex, and these are followed by a closer look at the input–output relations consistent with the theory.

§1. OUTLINES

The axons of the Purkinje cells form the only output from the cortex of the cerebellum (see Fig. 1); and these cells are driven by two essentially different kinds of input, one direct, the other indirect. The first is the climbing fibre input, and the second the mossy fibres, whose influence on the Purkinje cells may be complicated.

The inferior olive is the only known source of climbing fibres: every cell in the inferior olivary nuclei projects to the cerebellum, and every part of the cerebellum possesses climbing fibres (Eccles *et al.* 1967). Each of the rather small olivary cells sends out an axon which terminates in one climbing fibre on just one Purkinje cell: there are very few exceptions. The climbing fibre completely dominates the dendritic tree of the Purkinje cell, and its action has been shown to be powerfully excitatory (Eccles *et al.* 1967). Thus every olivary cell has a unique representational cell in the cerebellum which can be acted upon by all the influences mediated by the parallel fibres. In the present theory, it is suggested that each olivary cell corresponds to a 'piece of output' which it is necessary to have under control during movements. This 'piece of output' could take many forms: it might be a limb movement, or a fine digit movement, or an instruction to read vestibular output in a particular way to set up an appropriate control loop. Such 'pieces of output' will be called *elemental movements*; and each olivary cell may for the moment be supposed to correspond to one elemental movement in the sense that it is driven by an instruction for that movement to take place.

It is imagined that the olivary dictionary of elemental movements is complete: that is, every possible action can be represented as an ordered pattern of elemental movements each of which has a special olivary cell. Every action therefore has a defining representation as a sequence of firing patterns in the olive.

The final assumption, which relates the olivo-cerebellar system to the execution of motor actions, is that the nervous system has a way of converting the (inhibitory) output of a Purkinje cell into an instruction which provokes the precise movement to which its uniquely related olivary cell responds.

It will be argued that the reason for the special and in a sense substitutive relationship between a cell of the inferior olive and a Purkinje cell of the cerebellum is that the Purkinje cell can learn all the 'situations' in which the olive cell movement is required, and later, when such a situation occurs again, can implement that movement itself. If this were true of enough Purkinje cells (at least one for every elemental movement), the cerebellum could learn to carry out any previously rehearsed action which the cerebrum chose to initiate, for as that action progressed, the context for the next part of it would form, would be recognized by the appropriate Purkinje cells, and these would turn on the next set of muscles, allowing further development of the action. In this way, each muscle would be turned on and off at the correct moment, and the action would be automatically performed.

Information defining the context for each Purkinje cell is provided by

the mossy fibre input: and to establish that the Purkinje cells can learn contexts in the appropriate way it is necessary to demonstrate that the mossy fibre-granule cell–Purkinje cell arrangement could operate as a pattern recognition device. The notion fundamental to this is that the mossy fibre-granule cell articulation is essentially a pattern separator. That is, it amplifies discrepancies between patterns that are rather similar, translating two overlapping collections of mossy fibres into bundles of parallel fibres that overlap proportionately much less, if at all. One Purkinje cell can be made to store different contexts quite reliably by facilitating the relevant parallel fibre–Purkinje cell synapses: and this will work as long as the Purkinje cell does not try to learn too much. Evidently, the cue for synaptic modification is that the relevant climbing fibre be also active, and it is this which leads to the modification hypothesis.

These ideas lead to the notion that a mossy fibre input has been learnt by a given Purkinje cell if, and only if, the input is transformed into impulses in a bundle of parallel fibres all of whose synapses with that Purkinje cell have been facilitated. Two crucial points now arise. First, the number of parallel fibres into which a mossy fibre input is translated increases very sharply with the number of active mossy fibres unless the threshold of the granule cells also increases. The number of patterns each Purkinje cell can learn depends on the number of synapses which are facilitated in each: so economy arguments suggest that the granule cell threshold should be controlled in a suitable way. An inhibitory interneurone could achieve this, and the Golgi cells are interpreted as fulfilling this role.

The second point is that although the effect of the Golgi cells is to decrease the variation in the amount of parallel fibre activity, such variation will still exist. Whether or not a Purkinje cell should respond to a given mossy fibre input cannot therefore be decided by a fixed threshold mechanism. The Purkinje cell threshold must vary directly with the number of active parallel fibres running through its dendritic tree, and its actual value must be such that the cell emits a signal when and only when all the active parallel fibres have facilitated synapses with its dendrite. The natural way to implement this is to allow the parallel fibres to drive an interneurone which inhibits the Purkinje cell: and it will be shown that the various stellate inhibitory cells can be associated with this function, although their dendritic and axonal distributions are at first sight unsuitable.

1.1. *Data*

The anatomical and physiological information used in this paper concerns the cerebellum of cat, and is mostly derived from Eccles *et al.* (1967). Facts which are well known will not usually be given a reference:

information which is less well known is given a page reference in Eccles *et al.* (1967) if it appears there; otherwise an external reference is given.

A diagram of the general cerebellar cortical structure appears in Fig. 1. The cortex has two types of afferent fibre, the climbing fibres (*Cl*) and the mossy fibres (*Mo*). Each climbing fibre makes extensive synaptic contact with the dendritic tree of a single Purkinje cell (*p*), and its effect there is powerfully excitatory. The axons of the Purkinje cells leave the cortex (they form the only cortical output) and synapse with cells of the cerebellar nuclei.
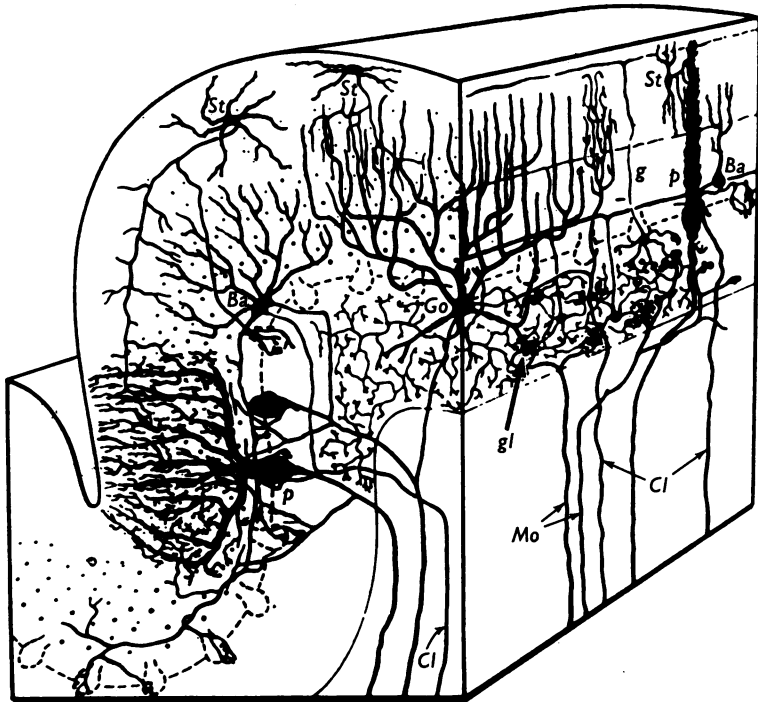

Fig. 1. Diagram of cerebellar cortex (from Eccles *et al.* 1967, Fig. 1). The afferents are the climbing fibres (*Cl*) and the mossy fibres (*Mo*). Each climbing fibre synapses with one Purkinje cell (*p*), and sends weak collaterals to other cells of the cortex. The mossy fibres synapse in the cerebellar glomeruli (*gl*) with the granule cells, whose axons (*g*) form the parallel fibres. The parallel fibres are excitatory and run longitudinally down the folium: they synapse with the Purkinje cells and with the various inhibitory interneurones, stellate (*St*), basket (*Ba*) and Golgi cells (*Go*). The stellate and basket cell axons synapse with the Purkinje cells, and the Golgi cell axons synapse in the glomeruli with the granule cells. As well as their ascending dendrites, the Golgi cells possess a system of descending dendrites, with which the mossy fibres synapse in the glomeruli. The Purkinje cell axons form the only output from the cortex, and give off many fine collaterals to the various inhibitory interneurones.

The second input, the mossy fibres, synapse in the cerebellar glomeruli (*gl*) with the granule cells. Each glomerulus contains one mossy fibre terminal (called a rosette), and dendrites (called claws) from many granule cells. The glomerulus thus achieves a considerable divergence, and each mossy fibre has many rosettes.

The axons of the granule cells rise (*g*) and become the parallel fibres, which synapse in particular with the Purkinje cells whose dendritic trees they cross. Where the granule cell axons (i.e. the parallel fibres) make synapses, they are excitatory.

The remaining cells of the cortex are inhibitory interneurones. The Golgi cells (*Go*) are large, and have two dendritic trees. The upper tree extends through the molecular layer, and is driven by the parallel fibres. The lower dendrites terminate in the glomeruli, and so are driven by the mossy fibres. The Golgi axon descends and ramifies profusely: it terminates in the glomeruli, thereby inhibiting the granule cells. Every glomerulus receives a Golgi axon, almost always from just one Golgi cell: and each Golgi cell sends an axon to all the glomeruli in its region of the cortex.

The other inhibitory neurones are stellate cells, the basket (*Ba*) and outer stellate (*St*) cells. These have dendrites in the molecular layer, and are driven by the parallel fibres. Both types of cell synapse exclusively with Purkinje cells, and are powerfully inhibitory.

Finally, the cortex contains various axon collaterals. The climbing fibres give off weak excitatory collaterals which make synapses with the inhibitory interneurones situated near the parent climbing fibre. The Purkinje cell axons give off collaterals which make weak inhibitory synapses with the cortical inhibitory interneurones, and perhaps also very weak inhibitory synapses with other Purkinje cells. These collaterals have a rather widespread ramification.

Behind this general structure lie some relatively fixed numerical relations. These all appear in Eccles *et al.* (1967), but are dispersed therein. It is therefore convenient to set them down here.

Each Purkinje cell has about 200,000 (spine) synapses with the parallel fibres crossing its dendritic tree, and almost every such parallel fibre makes a synaptic contact. The length of each parallel fibre is 2–3 mm ($1\frac{1}{2}$ mm each way), and in 1 mm down a folium, a parallel fibre passes about 150 Purkinje cells. Eccles *et al.* (1967) are certain each fibre makes at least 300 (of the possible 450) synaptic contacts with Purkinje cells, and think the true number is nearer 450. There is one Golgi cell per 9 or 10 Purkinje cells, and its axon synapses (in glomeruli) with all the granule cells in that region, i.e. around 4500. There are many granule cells ($2\cdot4 \times 10^6$ per mm^3 of granule cell layer), each with (usually) 3–5 dendrites (called claws): the average is 4·5 and the range 1–7. Each dendrite goes to one and only one

glomerulus, where it meets one mossy fibre rosette. It is, however, not alone: each glomerulus sees the termination of about 20 granule cell dendrites, possibly a Golgi cell descending dendrite, and certainly some Golgi axon terminals, all from the same Golgi cell. Within each folium, each mossy fibre forms 20–30 rosettes, giving a divergence of 1 mossy fibre to 400–600 granule cells within a folium. The mossy fibre often has branches running to other folia, and in Fig. 2 below one can count 44 rosettes on one fibre.

Just below the Purkinje cells are the Golgi cell bodies, and just above them are the basket cell bodies. There are 10–12 % more basket cells than Purkinje cells, and about the same number of outer stellate cells. Each basket cell axon runs for about 1 mm transversely, which is about the distance of 10 Purkinje cells. The basket axon is liable to form baskets round cells up to three away from its principal axis, so its influence is confined to a sort of box of Purkinje cells about 10 long and 7 across. The distribution of the outer stellate axons is similar except that it has a box about 9×7, since its axon only travels about 0·9 mm transfolially. The outer stellates inhabit the outer half of the molecular layer, and the basket cells the inner third. There are intermediate forms in the missing sixth. None of these cells has a dendritic tree as magnificent as that of the Purkinje cell, and Eccles *et al.* (1967) do not venture any comparative figures. Some outer stellates are small, with a local axonal distribution. A lot of the synapses of parallel fibres with this last group of cells are directly axo-dendritic, but all other parallel fibre synapses are via spines, though these are of different shapes on the different sorts of cell. Calculations based on slightly tenuous assumptions (in which Fig. 2 is an essential link) suggest that each Purkinje cell receives connexions from about 7000 mossy fibres: this will be explained in 3.1.

§2. CLIMBING FIBRES

The climbing fibre input has already been discussed at some length, and a formal statement of its part in the modification hypothesis will be made in 5.1. It is important to note that the fibre climbs like a creeper all over the dendritic tree of its chosen Purkinje cell, and forms synaptic contact almost everywhere. Each climbing fibre also sends terminals to other types of cell (basket, stellate and Golgi) in the vicinity of its Purkinje cell. These terminals seem to be excitatory, but only weakly so (Eccles *et al.* 1967, Table 1, p. 63). The climbing fibre collaterals and the Purkinje axon collaterals will be discussed together in 5.5.

§3. MOSSY FIBRES AND GRANULE CELLS

3.0. *The codon representation*

The synaptic arrangement of the mossy fibres and the granule cells may be regarded as a device to represent activity in a collection of mossy fibres by elements each of which corresponds to a small subset of active mossy fibres. It is convenient to introduce the following terms: a *codon* is a subset of a collection of active mossy fibres. The representation of a mossy fibre input by a sample of such subsets is called the *codon representation* of that input: and a *codon cell* is a cell which is fired by a codon. The granule cells will be identified as codon cells, so these two terms will to some extent be interchangeable. The size of codon that can fire a given granule cell depends upon the threshold of that cell, and may vary: and the mossy fibres which synapse with the granule cell determine the codons which may fire that cell.

There are exactly

$$\binom{L}{R} = \frac{L!}{R!(L-R)!}$$

codons of size R associated with a collection of L active mossy fibres. If two mossy fibre inputs each involve activity in L fibres of which M were common to the two, the two inputs are said to *overlap* by W elements; and they may be expected to have some codons in common. In fact the number they share is precisely $\binom{W}{R}$. The ratio X of the number of shared codons to the number of codons each possesses is given by

$$X = \binom{W}{R} \bigg/ \binom{L}{R} = \frac{W(W-1)\ldots(W-R+1)}{L(L-1)\ldots(L-R+1)} \tag{1}$$

which tends to $(W/L)^R$ as W increases. The limiting values of X for relevant values of R appear in Table 1. It will be observed that the effect of the subset coding is to separate patterns, because similar inputs have markedly less similar codons.

TABLE 1. Overlap Table, i.e. values of $(W/L)^R$

(W/L)	$R = 2$	3	4	5
0·5	0·25	0·12	0·06	0·03
0·6	0·36	0·22	0·13·	0·08
0·7	0·49	0·34	0·24	0·17
0·8	0·64	0·51	0·41	0·33
0·9	0·81	0·73	0·66	0·59

The mossy fibre-granule cell relay effectively takes a sample of the codon distribution of an input: the sample is small enough to be manageable, but large enough for the input event to be recoverable from it with high probability.

3.1. *The mossy fibre–Purkinje cell convergence*

A knowledge of the number of mossy fibres which may influence a given Purkinje cell is a prerequisite of a discussion of the codon sampling statistics: and this number may be estimated as follows. Let **P** be an arbitrary but henceforth fixed Purkinje cell: and assume that 200,000 parallel fibres synapse with **P**. Each granule cell has (on average) 4·5 claws,
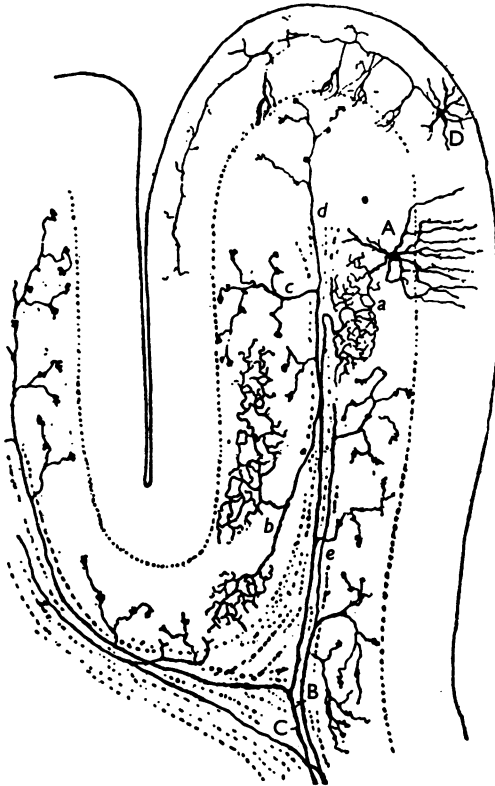

Fig. 2. Mossy fibres (B and C) terminating in two neighbouring folia (from Cajal, 1911, Fig. 41). The distribution of the terminals from each mossy fibre lies in the same plane as the axon of the basket cell (D). A is a Golgi cell.

so not more than 900,000 mossy fibres can influence **P** through the parallel fibres. Since the mossy fibre-granule cell divergence is 400–600 within a folium, the minimum figure for the mossy fibre–**P** convergence is 1500. It is apparent from Fig. 2 that the mossy fibre terminals occur in clumps of 4–10 rosettes, the average being 7 or 8, all of which might be expected to lead to granule cells most of which will contact any nearby Purkinje cell. If a mossy fibre leads to **P**, it may therefore expect to do so by 140–160 different paths (allowing for the divergence factor of 20 due to each

glomerulus). An average of 150 implies that about 6000 different mossy fibres lead to **P**. The edge effects will increase this figure, so 7000 would probably be a reasonable guess. This estimate will be used in the subsequent calculations.

3.2. *The assumption of randomness*

In the investigation that follows, it is assumed that the terminals of the 7000 mossy fibres are distributed randomly among the 200,000 granule cells leading to **P**. It is regrettable that no data exist to suggest a better model: and evidence will be produced (3.3.3 and 4.3) for the view that this assumption is actually false. Its value is that it enables computation whose results are at least illustrative: and one has the comfort of knowing that the capacity of a real cerebellum will anyway not be less than the result of calculations which assume a random distribution.

3.3. *The granule cell claws*

3.3.1. *Boundary conditions*. It will be assumed first that the claw arrangement of the granule cell dendrites is, as suggested by Eccles *et al.* (1967), a device to secure a high mossy fibre-granule cell divergence with minimal physical structure. But why do the granule cells have 4 or 5 claws and not more? These cells are extremely small and densely packed: and the parallel fibre synapses on **P** are extremely numerous. It is therefore reasonable to assume that the figure of 200,000 (or thereabouts) is the maximum physically realizable number of cells of this sort which can all send axons to **P**.

Secondly, it will be assumed that the synapses at the granule cells are not modifiable: that is, an excited mossy fibre will add a contribution to the excitatory post-synaptic potential (EPSP) of any granule cell with which it synapses, and this contribution has to be considered in determining whether or not that cell fires. This is justified below (3.3.3). Thirdly, the number of mossy fibres leading to **P** is of the order of 7000 in number: and fourthly, it is assumed that the system is to be used under conditions in which the number of active input fibres varies from around 20 to around 2000, if that is possible. There is clearly a need to allow considerable variation; some actions involve many more muscles and much more information from receptor organs than others. These figures are proposed as outer bounds, in the absence of any relevant evidence. Fortunately, it turns out not to matter crucially: the essential point is that the numbers are all nearer 0 than 7000 (on an arithmetical scale).

3.3.2. *Codon sampling*. The following rough model is used to calculate the number of granule cells per Purkinje cell that a given input can expect to stimulate. Suppose the number of active mossy fibres among the 7000

connected to the Purkinje cell **P** is L: then the number of possible codons is $\binom{L}{R}$, where R is the size of the codon. The number of codons which could be generated by the 7000 possible mossy fibres is $\binom{7000}{R}$; and if we assume that the granule cells have R claws and a threshold of R, then they represent a collection of 200,000 codons, supposed chosen randomly from the possible $\binom{7000}{R}$. Hence the number of granule cells per Purkinje cell stimulated by a given input of L active mossy fibres follows approximately binomial statistics with expectation

$$200{,}000 \binom{L}{R} \bigg/ \binom{7000}{R}. \tag{2}$$

The calculations that follow are concerned only with expectations: the numbers have in fact to be large enough for the distribution to be rather tightly clumped round the expectation. This is discussed in 5.2.3.

Suppose now that the granule cell has C claws and threshold $R \leqslant C$. That granule cell now has a catchment area of exactly $\binom{C}{R}$ codons of size R: and expression (2) becomes

$$200{,}000 \binom{C}{R} \binom{L}{R} \bigg/ \binom{7000}{R} \tag{3}$$

which is valid for expectations small compared with 200,000. (3) becomes (2) when $C = R$. The approximation (3) may be used, since it will be shown in 4.4 that situations will probably never occur in which the expectation is greater than 10,000. The values of (3) have been calculated for a selection of values of L, C, and R, and some of the results appear in Tables 2–4. Numbers greater than 20,000 have been replaced by an asterisk.

3.3.3. *Conclusions.* The conditions of 3.3.1 may be used to discover limits on the expected values of C and R. First, it is apparent from Table 4 that no codon size above 9 can ever be used when there are fewer than 13 claws per granule cell because too few granule cells would be activated. It is also evident and unsurprising that the maximum codon size used depends critically on the number of claws to each cell. Given this, the factor that will determine the number of claws to each cell will be economy of structure; and the relevant question is what is the least number of claws such that:

(i) The system is not swamped by large inputs: i.e. what is the least number of claws which still allows a small granule cell response to large inputs. Table 4 shows that a small response (less than 500) can be assured by 6 claws; so we expect to find at most 6.

(ii) The system remains sensitive to small inputs, if necessary by using

TABLE 2. Values of $200,000 \binom{C}{R}\binom{L}{R}/\binom{7000}{R}$: i.e. the number of the 200,000 granule cells synapsing with one Purkinje cell that a mossy fibre input involving L active fibres (out of 7000) can expect to stimulate. The granule cells have C claws, and threshold R

$$L = 20$$

R	C = 2	4	6	8	10	12
1	1134	2286	3429	4571	5714	6857
2	2	9	23	43	70	102
3	—	0	0	0	0	1
4	—	0	0	0	0	0
5	—	—	0	0	0	0
6	—	—	0	0	0	0
7	—	—	—	0	0	0
8	—	—	—	0	0	0
9	—	—	—	—	0	0
10	—	—	—	—	0	0
11	—	—	—	—	—	0
12	—	—	—	—	—	0

TABLE 3. Values of $200,000 \binom{C}{R}\binom{L}{R}/\binom{7000}{R}$: see legend to Table 2

$$L = 100$$

R	C = 2	4	6	8	10	12
1	5,714	11,429	17,143	*	*	*
2	40	242	606	1,132	1,819	2,667
3	—	2	11	32	68	125
4	—	0	0	1	2	4
5	—	—	0	0	0	0
6	—	—	0	0	0	0
7	—	—	—	0	0	0
8	—	—	—	0	0	0
9	—	—	—	—	0	0
10	—	—	—	—	0	0
11	—	—	—	—	—	0
12	—	—	—	—	—	0

TABLE 4. Values of $200,000 \binom{C}{R}\binom{L}{R}/\binom{7000}{R}$: see legend to Table 2

$$L = 2300$$

R	C = 2	4	6	8	10	12
1	*	*	*	*	*	*
2	*	*	*	*	*	*
3	—	*	*	*	*	*
4	—	2,327	*	*	*	*
5	—	—	4,582	*	*	*
6	—	—	251	7,016	*	*
7	—	—	—	657	9,862	*
8	—	—	—	27	1,213	13,339
9	—	—	—	—	88	1,943
10	—	—	—	—	3	191
11	—	—	—	—	—	11
12	—	—	—	—	—	0

an R less than C. The Table for information about this is Table 2, where for $L = 20$, we have to use $R = 1$ for all tabulated values of C, and it is not until $L = 100$ (see Table 3) that one can use $R = 2$ with $C = 6$. It

would therefore appear that to store inputs concerning fewer than 100 active mossy fibres, systems with $C = 6$ or less have to use codon size 1. This means a loss of discrimination between overlapping inputs of fewer than 100 active fibres. Provided however there are not many such small inputs, this will not be too serious. The number that must be kept negligible is the probability that a small, unlearnt mossy fibre input will occur all of whose active fibres have previously been involved in small learnt inputs.

This difficulty can to some extent be avoided if the mossy fibres which are active together in small input events have some tendency to grow near each other. The expected granule cell responses at codon sizes $R > 1$ will then be substantial at localized spots. This can be used, because it turns out that it is best to set the codon size on a local basis, rather than setting it uniformly over all the granule cells synapsing with a given Purkinje cell. The result for the animal will be greater reliability in its cerebellar responses, so mossy fibres which are correlated in this way could be drawn together by selection.

These arguments suggest that the arrangement of 4–5 claws per granule cell is consistent with structural economy and the conditions of 3.3.1. One point remains to be discussed: it is the assumption of 3.3.1 that the mossy fibre–granule cell synapses are unmodifiable. The most straightforward argument is this: every granule cell has a synapse with at least 300 Purkinje cells, each of which probably learns about 200 mossy fibre inputs (5.3). The chance that a given mossy fibre–granule cell synapse is used in none of these is extremely small (a generous estimate is 10^{-20}); whether or not it was initially facilitated, it almost certainly will be at some time. There is therefore no advantage in its being modifiable originally.

§4. THE GOLGI CELLS

4.0. *The need for variable codon size*

It became apparent in 3.3 that if the number of active parallel fibres was to remain reasonably small over quite large variation in the number of active mossy fibres, the thresholds of the granule cells had to vary appropriately. It will be shown (5.3) that the number of patterns a Purkinje cell can learn decreases sharply as the number of active parallel fibres involved in each increases. It is therefore essential to the efficient functioning of the system that the codon size should depend on the amount of mossy fibre activity.

4.1. *Requirements of a codon size regulator*

In the simple model containing one output cell **P**, 200,000 associated granule cells each making (possibly ineffective) synapses with **P**, and 7000

mossy fibres making contact in a random way with the granule cells, the task of a codon size regulator is in principle simple. It must count the number of these 7000 mossy fibres which are active, and set the threshold of the granule cells so that the following conditions are satisfied.

4.1.1. The number of active granule cells must be large enough to allow adequate representation of the mossy fibre input: that is, every active mossy fibre must with high probability be included in at least one codon.

This condition may be relaxed a little, since one factor on which the discriminatory power of the cerebellum depends is the accuracy with which the decision threshold at the Purkinje cell is set (5.2.3). There is no advantage in guaranteeing representation of the whole mossy fibre input if events slightly different from a learned event will anyway be responded to because of errors introduced later.

4.1.2. The number of active parallel fibres must exceed some lower bound N, where N will be taken as 500. This arises because the Purkinje cell threshold is not set directly from the parallel fibres with which it synapses, but from the results of sampling a number of different but closely related parallel fibres. The sampling is more reliable the more parallel fibres are active. This is explained in 5.2.3, where the figure of 500 is derived.

4.1.3. The codon size set for a particular mossy fibre input must depend only on that input; so that the same input is always translated into the same parallel fibres.

4.1.4. The codon size must be maximal, subject to conditions 4.1.1 to 4.1.3. This ensures that the number of modifiable synapses used for each learned event is minimal, and hence that the capacity is maximal (5.4).

It will be assumed that a signal in a mossy fibre is represented by a burst of impulses lasting many tens of milliseconds; and that a signal from a Purkinje cell is represented by a prolonged increase in its firing rate. This is discussed later (5.0); for the moment, it is needed only to justify the fifth condition.

4.1.5. The codon size regulating cell need not have set the granule cell threshold before the very first impulse in a signal arrives, but it must act very fast in response to such an impulse. It is essential that very little activity should be allowed into the parallel fibres while the granule cells are set at an inappropriately low threshold.

A mechanism to vary the threshold subject to these conditions could work in one of two ways: the threshold of the granule cells could be intrinsically high, and the mechanism provide excitation decreasing with increasing size of input; or the threshold could be low, and the mechanism provide inhibition increasing with increasing size of input.

4.2. *Properties of the Golgi cells*

The Golgi cells are inhibitory, can be driven by mossy fibres (through their descending dendrites) and synapse exclusively with granule cells. Further, they are particularly notable for the speed of their response (Eccles *et al.* 1967, p. 141). If the Golgi cells can be interpreted as a codon size setting device, it will therefore be as a mechanism of the second type described above.

There are, however, certain difficulties inherent in such an interpretation: first, each Golgi cell is driven by only a small number of the mossy fibre afferents to a single Purkinje cell, and sends an axon terminal to a relatively small number of granule cells; and second, the Golgi cells possess a large ascending dendrite system (Fig. 1), which on the present naive model is unexpected. The idea which the model lacks and which accounts for these various anomalies is the notion that Purkinje cells may share granule cells. Such sharing could clearly lead to great economies where two Purkinje cells needed codons from similar underlying subset distributions; but it is not obvious that sharing can be made to work, since two Purkinje cells may simultaneously require two different codon sizes.

4.3. *The effects of sharing granule cells among Purkinje cells*

If Purkinje cells are to be allowed to share granule cells, the assumption that the granule cell threshold should be constant over all cells synapsing with a given Purkinje cell must be abandoned. The most important single condition on the mossy fibre-granule cell transformation is (4.1.3) that it should be one-valued: a given mossy fibre input to a Purkinje cell should be carried there by parallel fibre activity which is determined by that input alone, and is independent of the simultaneous inputs to nearby Purkinje cells. This condition determines (in principle) the number and distribution of granule cells whose thresholds can be controlled together: for consider two adjacent Purkinje cells, P_1 and P_2. The collection of granule cells which synapse with P_1 but not with P_2 must be free to act as an independent unit, since it must be able to assume a threshold value different from the P_2 cells. If each parallel fibre is 3 mm long, and synapses with each of the 450 Purkinje cells that grow in 3 mm along a folium, the number of granule cells that synapse with a given Purkinje cell but not with its neighbour is about $200,000/450 = 444$.

The conclusion that may be drawn from these arguments is that the codon size should be set independently over blocks of about 450 granule cells. If this were done by an inhibitory cell, it should possess an axon distribution like that of the existing Golgi cells but limited to 450 granule cells, and a dendrite system like the descending Golgi dendrites: further there should be one such cell per Purkinje cell.

The fact that there are fewer and bigger Golgi cells than these arguments suggest must depend on certain information not incorporated in the model. This information concerns the distribution of the mossy fibre terminals which, if it were random and 4.1.3 were satisfied, would necessitate an arrangement near the expected one. In fact, one can see from Fig. 2 that the mossy fibres have a strong tendency to course transfolially, and in any case, given one mossy fibre rosette, there is a high probability that there will be another from the same fibre nearby. The effect of this, even apart from the considerations of 3.3.3, is to make nearby granule cells more related than they would be on the random hypothesis; and it is this which allows the larger axon and basal dendrite distribution which the Golgi cells are found to possess.

4.4. *The ascending Golgi dendrites*

The parallel fibre activity evoked by a mossy fibre input should be unique but perhaps more important even than that, it should involve rather few fibres, since the storage capacity of a Purkinje cell depends crucially on the number of parallel fibres active in each learned event (5.3). Some idea of the numbers of parallel fibres needed for various amounts of mossy fibre activity may be gained by using the simple random model. In Table 5, the expected number of active granule cells has been computed

TABLE 5. Possible codon size transitions (underlined); L is the number of active mossy fibres; R is the codon size

L	$R = 1$	2	3	4	5
100	12,857	323	4	0	0
300	*	2,929	109	2	0
500	*	8,148	507	15	0
700	*	15,979	1,395	60	1
900	*	. *	2,967	163	3
1,100	*	*	5,420	364	10
1,300	*	*	8,950	711	22
1,500	*	*	13,754	1,261	45

for inputs with L active mossy fibres, these L chosen at random from a population of size 7000. The calculation has been performed on the assumption that 100,000 granules have 4 claws, and 100,000 have 5: for a threshold of R, the approximation used was

$$\text{expected number} = 100,000 \left(\binom{5}{R} + \binom{4}{R} \right) \binom{L}{R} \bigg/ \binom{7000}{R} \qquad (4)$$

which is derived the same way as expression (3), and is valid only for answers small compared with 100,000. The codon size transition regions have been underlined. It will be observed that on this rather crude model, each input arouses between 500 and 9000 granule cells: (500 is the lower

bound justified in 5.2.3). If as many as 9000 are used, the capacity of each Purkinje cell will be drastically reduced. For large values of L, around 1000 (if such are ever used) the figure of 9000 is not unreasonable. Indeed some number of that order will be necessary to cover all the active incoming fibres. But if not, and it is questionable whether such large inputs ever occur, then to use so many would be wasteful. Provided that the total number of active parallel fibres is greater than 500, it is possible to use condition 4.1.1 to submit this number to an upper bound which depends on the amount of mossy fibre activity. For example, the number of parallel fibres active should only exceed 5000 if the number of active mossy fibres exceeds 500.

The upper dendritic tree of a Golgi cell may be interpreted as a mechanism to superimpose this upper bound; and it may be expected to work as follows. A mossy fibre signal arrives which may be quite different from what was going on before. The descending Golgi dendrites sample it and quickly set new thresholds at the relevant granule cells: this setting amounts to a first guess based on local sampling. Rather a long time later, the signals appear in the parallel fibres, and the Golgi cell, by examining the activity in a large number of these, can tell whether or not its initial assessment was the most economical solution. If it was, its behaviour should not alter: if not it should; but this will always entail shifting to a higher codon size. One cannot say that the local or global sample will always give the best solution; for example, it might happen that the mossy fibre input is sufficiently localized that it can support a high codon size for just one or two Golgi cells.

In general, therefore, a Golgi cell should be driven by that dendritic system from which it receives most excitation. This suggests that the upper and lower dendritic fields should have rather a peculiar relationship. The synaptic influences among the upper dendrites should summate, and so should the effects among the lower ones: but the summed contributions should interact so that the output from the cell is driven by the maximum of the two, not the sum. There is no firm evidence to support this prediction, but Eccles *et al.* (1967) mention that the two dendritic fields are probably too far apart to allow summation.

A proper investigation of the action of the Golgi cells would be difficult for two reasons. First, one cannot use a random model for the way the granule cells are distributed over the possible subsets of the mossy fibres, for as well as the objections of 4.3, it is likely that mossy fibres whose activities are correlated will grow near one another. This is because input events would then tend to need fewer granule cells to cover them, and could therefore be more economically stored. Secondly, an analytic model of the relationships between neighbouring Golgi cells under various input

conditions needs unrealistic simplification before it can be handled. The correct approach is probably to use a simulation programme; and the kind of result to which it will lead is that an action A, learnt in isolation, may have to be relearnt to some extent if it is to be performed immediately after some closely related action B. This will arise because the hangover in the parallel fibres of action B could cause temporarily different codon sizes in certain Golgi cell blocks. The parallel fibres for A in this situation are slightly different from those for A performed in isolation. In other words, the price of economy is probably a not too serious loss of uniqueness for the mossy fibre–parallel fibre transformation.

4.5. *The Golgi cell afferent synapses*

It will be clear that within the present theory, no advantage would be gained by having the mossy fibre–Golgi cell synapses modifiable: but it is not so clear whether this is also true of the parallel fibre–Golgi cell synapses. Although there is no very simple way in which it would be useful to have these synapses modifiable, it is conceivable that there might be fringe benefits. Suppose, for example, that activity in a particular set of mossy fibres always preceded a large volley: then such advance information could be used by the Golgi cell if the conditions under which modification took place were arranged suitably.

On the other hand, modifying a synapse on a Golgi cell implies that the parallel fibre has a special relationship with the granule cells below that Golgi cell. Leaving aside the case that it came from one of them (not a special relationship of the relevant kind) there is no reason why, even if such a relationship were to hold over a number of inputs, it should hold over a majority, since one Golgi cell can expect to serve a huge number of different facilitated responses. And, in contrast to the Purkinje cells, there are no inhibitory cells of any power acting upon the Golgi cell, so there is no mechanism for deciding whether or not a majority of the currently active fibres have or have had such a special relationship. (The absolute size of a 'majority' is variable: so the Golgi cells would need a variable threshold to make such a decision, for the same reasons as do the Purkinje cells.) This argument suggests rather strongly that these synapses are not modifiable.

The other afferent Golgi synapses come from the Purkinje cell collaterals and the climbing fibre collaterals: these will be discussed in 5.5.

§5. THE PURKINJE CELLS

5.0. *The Purkinje cell output*

The main branch of a Purkinje cell axon goes to one of the cerebellar nuclei and forms the only output from the cortex: its effect in the nucleus is inhibitory. It is an assumption of the present theory that the central nervous system has a means of converting a signal in a Purkinje cell axon into an order to provoke the elemental movement to which its corresponding olivary cell responds. The inhibitory nature of the Purkinje cell output suggests that it may require a positive effort to read from the cerebellum, since excitation must be fed in somewhere to the effector circuit. This would be useful (though not essential) if it were required that cerebellar output should often be ignored: and indeed it is likely that such occasions will frequently occur during waking life and possibly also during sleep. The fact that the cortical output is inhibitory can therefore be interpreted as a convenience for easy ignoring, though this is neither the only nor a necessary view.

The second point arises from the fact that Purkinje cells have a high resting discharge of 20–50 impulses/sec. (Eccles *et al.* 1967, p. 306). It was assumed above (4.1.5) that a signal in a mossy fibre was represented by a burst of impulses: and the codon size setting function of the Golgi cells depended upon this. It is a necessary consequence that efferent cortical signals should also be represented by a train of impulses rather than a single one, since the delays involved in turning on the inhibitory interneurones could make the initial response of a Purkinje cell to a mossy fibre input inappropriate. This may occur frequently, and would conveniently be hidden by a high resting discharge. Purkinje cells can sustain high rates of firing (greater than 400/sec, according to Eccles *et al.* 1967, p. 308): it is therefore reasonable to assume that a signal in a Purkinje cell axon is represented by a large increase in the firing rate, and that the effector systems are only sensitive to such messages. This assumption would have to be made for almost any theory of the cortex, since the Purkinje cells form the only output.

The input–output relations for the cortex as a whole receive attention in §7, and the Purkinje axon collaterals in 5.5.

5.1. *The hypothesis of modifiable synapses*

The fundamental hypothesis for the mechanism of the change of effectiveness of a parallel fibre–Purkinje cell synapse is that *if a parallel fibre is active at about the same time as the climbing fibre to a Purkinje cell with which that parallel fibre makes synaptic contact, then the efficacy of that synapse is*

increased towards some fixed maximum value. ('At about the same time' is an intentionally inexact phrase: the period of sensitivity needs to be something like 50–100 msec.)

If this hypothesis is true, it may have implications about the physiological conditions for synaptic modification. The most striking fact about the climbing fibres is that they extend over the whole Purkinje dendritic tree. Two of the possible reasons for this seem plausible: first, that the climbing fibre releases some sort of 'change' factor which modifies the active synapses; or second, that the fundamental condition for modification is simultaneous pre- and post-synaptic depolarization. Hebb (1949) suggested that the nervous system might contain synapses with modification conditions of the second sort.

The other and rather dangerous place one might look for implications of the modification hypothesis is in the comparison of electron-micrographs of cells supposed to have modifiable synapses with those supposed not to. This will not be attempted, but it may be relevant that the Purkinje cells seem to be the only ones in the cerebellum whose dendrites carry the characteristic tubular system which terminates 'abruptly' at the base of the spines (Eccles *et al.* 1967, p. 9).

5.2. *Simplifying assumptions*

The calculation of the learning capacity of a single Purkinje cell requires that certain simplifying assumptions be made.

5.2.1. It will be assumed that a synapse is either totally modified or totally unmodified: and that stimulation of a totally unmodified synapse has no effect on the post-synaptic membrane.

This is equivalent to allowing modification to increase synaptic efficacy from some fixed minimum to some fixed maximum value in one step: since the two situations can be identified by subtracting any 'ground' excitation of an unmodified synapse. Such a subtraction has a linear dependence on the number of parallel fibres active at any moment, and could easily be performed by an unmodifiable inhibitory interneurone such as the basket or outer stellate cells. This may indeed be one of the functions of these cells: it is a matter of no importance to the present theory, since such an effect would be constant throughout the life of the cerebellum. The phrase 'in one step' is merely a conceptual convenience: the matter will be discussed in § 7.

5.2.2. Secondly, it will be assumed that each learned event occupies a set of parallel fibres which may be regarded as having been chosen at random from the 200,000 which synapse with the Purkinje cell. This assumption can be justified on the grounds first that any estimation of capacity arrived at by using it is likely to be too low; and, secondly, that

almost any other assumption would involve a computational effort out of all proportion either to the probable truth or to the value of any results thereby achieved. The assumption is likely to be false for two reasons: first, the mossy fibre–granule cell relay is probably not randomly constructed; and, secondly, the learned events are unlikely to have a random structure over the space of possible mossy fibre inputs. Some types of structure on the learned inputs will positively confuse the system into giving false responses.

This topic will receive a full and more general analysis in a later paper: two remarks however are not out of place here. If there are many learned mossy fibre inputs which all overlap each other by a considerable amount, the cerebellum may not be able to discriminate inputs which have been learned from inputs which have not when an unlearned input has much overlapping with learned inputs. The fact that no granule cells have more than 7 claws introduces an absolute upper bound to the discriminatory power of the cerebellum: and, when this is inadequate, control must revert to the cerebrum. It should also be noted that any subset of a learned mossy fibre input will behave as a learned input if it causes the same codon size range to be selected as did the learned input. In the full model, the condition is more restrictive in that the codon size range must be the same for most of the 150 or so Golgi cells concerned.

5.2.3. The maximum desirable number of facilitated synapses on any one Purkinje cell will be taken as 140,000, and the minimum number of parallel fibres active in any learned event as 500.

These figures are related by the way the Purkinje cell threshold is set (see 6.1). It turns out that the most economical way of doing this is by sampling a population of parallel fibres closely related to and including the ones passing through the Purkinje cell dendritic tree. Let $T(E)$ be the threshold set in response to the stimulation of $M(E)$ parallel fibres by the mossy fibre input E (regarded as an input to a particular Purkinje cell **P**). Let f be the fraction of the (200,000) parallel fibre synapses which have been facilitated at **P**. If E has been learnt, all $M(E)$ of the active parallel fibres will have facilitated synapses at **P**. Hence if E is to be recognized as learnt, $T(E) \leqslant M(E)$ (i).

If E is not a learned event, and $F(E)$ is the number of the active parallel fibres which have facilitated synapses at **P**, then E will be ignored only if $T(E) > F(E)$ (ii).

If recognition is a reliable process, both (i) and (ii) must be true with high probability.

The randomness assumption 5.2.2 allows us to assume that $F(E)$, taken over events E with constant $M(E)$, has a binomial distribution with expectation $fM(E)$. It is unlikely that T (taken over the same event population)

has worse than a binomial distribution, since the binomial assumption is equivalent to regarding the value of T as being set on the basis of a measurement of the number of active mossy fibres involved in E. If one assumes T has a binomial distribution, conditions (i) and (ii) are satisfied with probability rather greater than 0·99 if $M \geqslant 500$, $f = 0·7$. These values for f and the minimum value of M will be used despite the rather low confidence level because (*a*) the threshold setting mechanism is certain to be better than a binomial process, and (*b*) few input events will use the minimum number of parallel fibres allowed.

5.3. *The storage capacity of a Purkinje cell*

The capacity of a Purkinje cell may be calculated very simply from the assumptions 5.2. Suppose the fraction of facilitated parallel fibre synapses is 0·7, and each learned event occupies n parallel fibres. Then x, the expected number of events which may be learned before the total proportion of synapses used exceeds 0·7, is the largest integer for which

$$(1 - n/200,000)^x > 0·3.$$

x has been computed for various values of n, and the results appear in Table 6. It will be seen that the advantage of having a small number of fibres active in each learned event is an enormous increase in capacity: the Golgi cell arrangement of local as well as global constraints on the codon size begins to make good sense. If the minimum number of parallel fibres active in learned event is 500, the average number of responses stored by each Purkinje cell is probably in excess of 200.

TABLE 6. x is the number of events each occupying n parallel fibres that can be learned by one Purkinje cell, i.e. x is the largest integer for which

$$\left(1 - \frac{n}{200,000}\right)^x > 0·3$$

n	500	1,000	2,000	5,000	10,000	20,000
x	480	240	119	47	23	11

5.4. *The Purkinje cell threshold*

The inhibition of the basket and outer stellate cells can be a powerful influence on the behaviour of a Purkinje cell. The present theory requires that the Purkinje cell should fire if and only if more than a proportion p of the active parallel fibres have facilitated synapses with it, where p is close to 1. It is proposed that the purpose of these stellate cells is to provide the appropriate inhibition, and that their peculiar axonal distribution is a device to secure an economy of dendrite by a factor of up to 20 (see §6).

It is made possible by the distribution of the mossy fibre terminals beneath them.

If M of the parallel fibres synapsing with a particular Purkinje cell are active, and Ms of these have been facilitated, then the cell must fire (or produce a burst of firing) if and only if $Ms \geqslant Mp$. The Purkinje cell thus has the superficially simple task of summing Ms (represented by excitation) and Mp (represented by inhibition). If, however, one reflects upon the enormous expanse of the Purkinje cell dendritic tree, it becomes apparent that to arrange such a summation might not be an easy problem of dendritic engineering. The example in Plate 1 makes it difficult to imagine how the junction of a spiny branchlet with the rest of the dendrite could carry accurate information about the number of active spines if this were large, for if the 100 nearest the junction were active, it is hard to see how (say) 10 at the end of the branchlet could make much difference, at least on any simple view of dendritic function. Such a system can only provide accurate summation for numbers of active synapses rather small compared with the total population.

This overload effect can be overcome locally if the number of active fibres is kept small: but it is bound to recur on a larger scale unless the numbers are kept very small. A further trick seems necessary, and the right one is probably to do the subtraction piecemeal: add up the outer contributions to Ms, subtract the outer Mp, and transmit the result to be added to the contribution to Ms of the next region. This is the only way of subtracting B from A with large A and B but small $(A - B)$ without ever handling large numbers.

The distribution of the axon terminals of the basket and outer stellate cells is peculiarly well suited to this interpretation. The outer stellate cells effectively sample the activity in the outer half of the molecular layer and send their (inhibitory) contribution to $-Mp$ to a region quite high up the dendritic tree of the Purkinje cell. The basket cells sample about the inner third, sending their contribution to the soma; and the intermediate cells perform an intermediate task. The basket cell action represents the last stage in computing $(Ms - Mp)$, and one may assume that the numbers are then small enough for the coding from dendrite to soma to be adequate.

Interpretation of the function of simple summation within any reasonable theory of dendrites would be made easier by two hypotheses: first that the number M of active parallel fibres was both small and reasonably constant; and secondly, that the excitation due to a facilitated synapse differed very little between synapses. In view of the existing Golgi cell arrangement and the great increase in capacity which is a consequence of having M small, there are strong reasons why the first hypothesis should be true. And the

second may be a consequence of the fact that these synapses are all spine synapses, which possess a definite morphological uniformity.

5.5. *The Purkinje axon collaterals and the climbing fibre collaterals*

The axons of the Purkinje cells give off numerous fine collaterals which form two plexuses. The infraganglionic plexus lies below the Purkinje cell bodies, and its fibres run in a predominantly transverse direction. The supraganglionic plexus, which is fed both directly and by branches of fibres rising from the infraganglionic plexus, lies above the Purkinje cell bodies, and its fibres run in a predominantly longitudinal direction (Eccles *et al.* 1967, p. 178). Not a great deal is known about the distribution of the Purkinje cell collaterals, but it seems that at least in the vermis the spread of the collaterals in the longitudinal direction is small, whereas in the transverse direction it may be quite large; and the longer collaterals tend to join points of cortex to their corresponding contralateral points. (See Eccles *et al.* 1967, pp. 178ff., for a discussion and references.)

These collaterals have weak inhibitory synapses with basket and Golgi cells, and perhaps also very weak inhibitory synapses with other Purkinje cells (Eccles *et al.* 1967, pp. 184ff.). Their effect through the basket cells is to release Purkinje cells from inhibition, but their influence through the Golgi cells is more complicated. It is likely that this influence will ultimately be excitatory at a given Purkinje cell **P** only if most of the granule cells thereby released from Golgi inhibition have modified synapses at **P**: and this will be true only if **P** has already learned a number of mossy fibre inputs all quite similar to the current input.

The only obviously reasonable interpretation of the effect of these collaterals is that they tend to excite the Purkinje cells in the cortex to which they distribute; and in certain circumstances can loosen the discrimination exercised by those cells. The fact that a Purkinje cell P_1 has just fired may be relevant in a borderline firing decision for **P** if **P** and P_1 lie in closely related pieces of cortex: and the Purkinje axon collaterals provide a suitable means of distributing this information. They can help **P** overcome inhibition due, perhaps, to an unlearned mossy fibre input which it has previously received, or they can make **P** more likely to accept the current input even though it may not be exactly one which has been learned.

This view is not entirely satisfying, but it does provide an interpretation of the climbing fibre collaterals. These make weak excitatory synapses with the inhibitory interneurones of the cortex (Eccles *et al.* 1967, Table 1, p. 63), and perhaps with Purkinje cells. Their distribution is more local than that of the Purkinje axon collaterals (Eccles *et al.* 1967, p. 215), but their effect locally probably roughly balances them. Hence it could be

argued that when a climbing fibre is active, that is when synaptic modification is taking place, the effect of Purkinje axon collaterals is at least partly annulled, and so something nearer a true representation of the mossy fibre input is stored.

§6. THE OUTER STELLATE AND BASKET CELLS

6.0. *Justification of their joint treatment*

The outer stellate and basket cells will be taken together under the general heading of stellate cells for the following reasons.

6.0.1. They are both inhibitory.

6.0.2. They both send axons to the Purkinje cells only.

6.0.3. They are both driven mainly by parallel fibres, and have analogous dendritic fields, the outer stellates being further out in the molecular layer.

6.0.4. They have very similar axon distributions; the outer cells synapse further up the Purkinje cell dendritic tree, and reach a little less far across the folium than the inner ones.

6.0.5. There exist many intermediate forms.

The discussion will also include in a general way the apparently rather weak 'on-beam' outer stellate cells whose axons terminate locally, though these will receive special mention.

6.1. *The function of the stellate cells*

The stellate cells together have the task of controlling the threshold of the Purkinje cells: they are powerful, and have to be, since if the overload ideas 5.4 are correct they have to be able to contain almost the maximal excitation that parallel fibre activity can evoke in the Purkinje cell dendritic tree. (This, it was argued, is achieved long before all the parallel fibres are active.) The quantitative relations between the number of parallel fibres active and the strength of the inhibition necessary have been discussed in 5.2.3, and reasons for the distribution of the terminals on the Purkinje dendritic trees have been proposed in 5.4. It remains only to sort out two points: the size, shape and position of the dendritic tree, and the distribution of the Purkinje cells to which the stellate cells send axon terminals.

If one naively set about constructing a threshold-setting cell to perform the function required by the present theory, one would propose one inhibitory cell per one or two Purkinje cells. Its axon would synapse with just the one (or two adjacent) Purkinje cells, and its dendritic field would at least be very close to that of its corresponding Purkinje cell. If there were such cells, however, their dendrites would have to be not only very

close to those of the relevant Purkinje cell, but also very nearly as extensive: this would be necessary in order to obtain a reliable measurement of the usually sparse parallel fibre activity.

The reason why the stellate cells are not arranged like this is that since such a dendritic tree would necessarily take up roughly as much room as does a Purkinje cell dendrite, the number of Purkinje cells that could be packed in any given length of folium would be about halved. The key to the success of the existing solution is that the rosettes of each mossy fibre are numerous and on the whole distributed transfolially in the granular layer. The actual mossy fibres that drive the cortex therefore change quite slowly across a folium, and they can be watched efficiently and economically by sampling the parallel fibre activity across it (Fig. 2).

There is no quantitative evidence available from which one might investigate the tenability of this hypothesis: one can only estimate the economies to which the proposed sampling technique may lead. Each Purkinje cell receives inhibition from about 40 stellate cells: the inhibition to the Purkinje cell is therefore driven by a dendritic field about 40 times as large as that of a single stellate cell. If these 40 were distributed randomly just next to the Purkinje cell, a good sample (approximately $1 - 1/e^2 = 0.86$) would be obtained if each tree had even 1/20 of the synapses that a Purkinje cell has. If the mossy fibre distribution alters slowly (which it has to do anyway for the system to work), the saving in dendrite could therefore be a factor of up to 20; and, in practice, the sampling is certainly not random.

6.2. *The stellate cells with local axonal distribution*

It is convenient to complete the review of the cells of the cortex with some remarks about the time courses of the excitatory and inhibitory synaptic actions. It is evident that the time course of transmitter action at a Purkinje cell is the ultimate factor determining the temporal extent of the influence on that cell of information from that fibre.

At a normal sort of synapse, such influence would not be expected to continue more than 20 msec after activity in the afferent axon had ceased: but so short a period would seem inappropriate for real-time analysis of events with characteristic times rarely less than 100 msec. The observed time course at a parallel fibre–Purkinje cell synapse is of the order of 100 msec, and Eccles *et al.* (1967) mention (p. 70) that this may be one function of spines. Now the connectivity of cerebellar cortex is such that the onset of the various post-synaptic effects at a Purkinje cell due to a mossy fibre signal is likely to be both slow and patchy: the various components arrive along paths with different latencies, and there may be build-up effects in the synapses themselves. Similar factors will affect the way the post-

synaptic effects decay at the end of a mossy fibre signal. All these effects can only be disruptive, at least as far as the present theory is concerned. The fact that the various effector circuits in the rest of the nervous system are geared only to recognizing bursts from Purkinje cells will minimize the effect of any stray impulses which might for a number of reasons leak out: but it is conceivable that the effects of an input during a 'turning on' or 'turning off' period could cause a false response from a Purkinje cell, and that response could last up to 20 msec.

In order that false outputs of this sort should not occur, it is necessary that the build up of inhibition at a Purkinje cell should occur faster than the build up of excitation, and that the IPSP should last longer than the EPSP. The latter is an observed phenomenon, with IPSP time courses up to 500 msec, EPSP ones up to about 100 msec; and it is possible that one function of the 'on-beam' stellate cells is to ensure the former. These cells have a local axonal distribution, so their axons are relatively very short; and many of their synapses with the parallel fibres are direct (i.e. not spine synapses). The first factor must, and the second may favour a fast production of IPSP at the Purkinje cell dendrite, and this IPSP could well arrive early enough to counteract the initial build up of EPSP from the Purkinje spines. The IPSP induced by these cells is weak, but by the time the Purkinje cell is turned on to any appreciable extent, the other stellate cells will also be active. It is therefore proposed that the weak on-beam stellate cells be interpreted as a device to prevent a false initial response by the Purkinje cell.

§7. CEREBELLAR INPUT–OUTPUT RELATIONS

There are two main types of cerebellar input–output relation which are compatible with the present cortical theory and they are described separately.

7.1. *Learned movements*

The first possibility is the one suggested in §1, and concerns the learning of that particular sort of motor skill which may be described as a movement. During learning, the cerebrum organizes the movement, and in so doing, causes the appropriate olivary cells to fire in a particular sequence. This causes the Purkinje cells to learn the contexts within which their corresponding elemental movements are required, so that next time such a context occurs the mossy fibre activity stimulates the Purkinje cell, which evokes the relevant elemental movement.

This scheme imposes severe restraints upon the nature of the stimulus that may drive an olivary cell: indeed, almost the only permissible case is that in which each olivary cell is driven by a collateral of a cerebral command

fibre for some elemental movement. This statement may be justified by the
following argument. During execution of a learned movement, the mossy
fibre activity is responsible for the initiation of the various elemental
movements: and it is therefore essential that, during learning, the Purkinje
cell is associated with the context occurring just *before* its elemental
movement. The present theory suggests that the granule cells and Golgi
cells together provide extremely effective pattern discrimination: so the
mossy fibre activity must be virtually the same during cerebellar execution
of a movement as it was while that movement was being learnt. Hence, for
the cerebellum to be able to learn a movement in which the contexts
change rapidly, the olivary activity during learning has to be driven by
impulses effectively synchronized with the commands. This conclusion can
only be avoided in one of the two following ways: either some delay is
specially introduced into the mossy fibre afferents, or the olivary cells are
driven by the elemental movement just preceding the current one. The
first assumption is unlikely on grounds of efficiency, and the second would
require a probably unacceptable number of olive cell-Purkinje cell pairs,
one for each sequence of two elemental movements.

The above argument, however, cannot be applied to those situations
where the contexts are changing very slowly: and in such cases it is at least
logically possible for an olivary cell to be driven by a signal which was
slightly later than the command signal during learning, since the relevant .
context will scarcely have altered. It is therefore not impossible for an
olivary cell to be driven by a receptor which is sensitive to the movement
initiated by its corresponding Purkinje cell: although, if the contexts do
change slowly, a context driven system will not reproduce the timings of
the stages in a movement at all accurately, and so cerebellar learning will
anyway be rather bad.

It can therefore be concluded that an olive cell–Purkinje cell pair, whose
olive cell is driven by a receptor, is unlikely to be used for learning motor
skills involving much movement. It is however well known that the inferior
olive is divided into two portions, one driven by descending fibres (Walberg,
1954) and one by ascending fibres (Brodal 1954). Further, it is known that
at least some cells in the 'ascending' or 'spinal' part of the olive are
driven by receptors (Armstrong, Eccles, Harvey & Matthews, 1968), and
these authors also demonstrate the convergence at some cells of impulses
from receptors of quite different types.

If the present cortical theory is correct, and the cerebellum does learn
motor skills, there is only one situation in which it is not absurd to drive
the olivary cells by receptors rather than by cerebral command fibre
collaterals, and that is when the cerebellum is required to carry out an
action in a different language from that in which the cerebrum originally

set it up. This condition is likely to be fulfilled in the cerebellar control of balance or posture, where one may reasonably expect the cerebrum to deal in a language oriented towards the problem of changing postures, while the cerebellum is concerned primarily with maintaining an achieved posture. It is to this kind of control that the second form of input–output relation is particularly well suited, and it will be discussed in detail in 7.2.

If it is assumed that such situations are best dealt with by the methods described in 7.2, the following conclusion may be drawn. Where the cerebellum is required to learn a motor skill consisting of a movement, the cells of the inferior olive should be driven by the equivalent of a collateral of the cerebral command fibre for a particular elemental movement; and the Purkinje cell corresponding to that olivary cell should be able to provoke that same elemental movement. The particular elemental movement associated with an olivary cell–Purkinje cell pair need not be fixed, but it presumably is: and the elemental movements associated with this kind of input–output relation are probably mostly small movements.

To complete the study of this kind of input–output relation, four further points must be discussed. The first concerns a possible variant in the way information is read out of the cerebellum. It was assumed in §5 that the level of inhibition at a Purkinje cell was generally rather low, and that mossy fibre activity involved in a learned context was enough to produce a signal in the Purkinje cell axon. There is another possibility, in which the level of inhibition at a Purkinje cell is generally rather high, and the rest of the brain decides whether the current context has been learned by observing the results of a climbing fibre impulse. If the mossy fibre input has been learned, the Purkinje cell gives a large response; if not, it gives a small one and the effector circuits respond accordingly. This form of output may be described as *inhibition sampling*, and has effectively been suggested by Eccles *et al.* (1967, p. 177), though not in the context of modifiable synapses.

The second point concerns the command circuit used by the cerebrum while setting up a movement. It is possible that the olivary cells are literally driven by collaterals of the cerebral command fibres: but it is also possible that the command circuit actually is the cortico-olivo–Purkinje cell–effector circuit path. This hypothesis involves no difficulties and is especially attractive if Purkinje cell output is obtained by inhibition sampling: for this could then easily be achieved by uniform weak descending activity to the inferior olive, arriving by the same pathways as are used for the cerebral organization of movements. One extra hypothesis is also needed if this system is postulated, namely, that the mechanism of synaptic modification at Purkinje cells is sensitive only to intense climbing fibre activity.

The third question arises from the possibility that synaptic modification may be subject to gradual decay. This might be necessary, in view of the limited learning capacity of a Purkinje cell: and one might imagine that repetition of a movement carried out even under cerebellar control should have some reinforcing effect. If a Purkinje cell command were somehow fed back to excite the relevant climbing fibre, a reinforcing effect would certainly be obtained, but lack of this feed-back would not rule out the possibility that a reinforcing effect exists, since this depends on the details of the synaptic modification mechanism.

The final point to be raised in this discussion of the cerebellar control of movements is the question of the speed with which such movements are executed. There is no reason why a context dependent system should not be run at different speeds, and if the extra postulate were made of some general intensity control acting uniformly over the effector circuits, a movement learnt at one speed could be performed at another. (This idea would fit nicely with the suggestion made above that during cerebellar control of a movement, the olive receives uniform weak descending activity, for the strength of the Purkinje cell output would then depend upon the strength of this uniform activity.) It is, however, likely that if the time course of a movement were changed substantially, some relearning would be necessary.

7.2. *Learned conditional reflexes*

The explanation of the second type of input–output relation compatible with the cortical theory is much simplified by the introduction of a new idea, which extends the classical notion of reflex.

Definition. A *conditional reflex* is a reflex which operates when, and only when, certain conditions outside the reflex arc are satisfied. These conditions are the *context* of the conditional reflex, and a *learned conditional reflex* is a conditional reflex whose context is learned.

An ability to acquire learned conditional reflexes would make the task of maintaining balance and posture very much easier for the nervous system. For example, consider the problems which confront a child as he learns to stand. It would greatly aid him if he could form a reflex circuit which connected a vestibular signal indicating some imbalance directly to an order for the appropriate compensating movement: this, however, could not be a true reflex since the child will not always wish to stand. The appropriate form of control is a conditional reflex whose context is the state of standing, and which therefore only operates while the child is standing. In order to suspend the standing reflexes, the child has only to disrupt the 'standing' context, and this could be done, for example, by his suddenly wishing to stand no longer.

A particular olivary cell–Purkinje cell pair may be interpreted as a storage unit for a conditional reflex if and only if the circuit environment $\rightarrow$ receptor $\rightarrow$ olivary cell $\rightarrow$ Purkinje cell $\rightarrow$ effector $\rightarrow$ environment is a stabilizing negative feed-back loop when activated by a learned mossy fibre input. The context represented by the learned mossy fibre input is the context of the conditional reflex. Explicitly, the conditions for storage of a conditional reflex are as follows:

(i) Output is obtained from the Purkinje cell by the inhibition sampling method (described in 7.1): that is, the level of inhibition is generally high, so that climbing fibre signals are only transmitted when the mossy fibre input is one that has been learned.

(ii) The olivary cell is driven by receptors whose stimulation is reduced (in any learned context) by the results of stimulating the corresponding Purkinje cell.

The learning of a context will arise if the combination of olivary cell firing and that particular context is a frequent one, as it would be, for example, while the child (under cerebral control) was 'learning' to stand. Once the context is learned, the reflex automatically becomes operative when it is required.

There is no reason why a particular olivary cell should not be driven by more than one kind of receptor, though receptors must be connected to Purkinje cell units whose activity reduces the stimulations they receive: the inhibitory nature of the Purkinje cell output may help to arrange this. The receptors connected to a given olivary cell have to be rather carefully chosen, but their number is limited only by the learning capacity of the corresponding Purkinje cell.

It is proposed that most cerebellar functions associated with maintaining balance and posture are carried out by forming the appropriate learned conditional reflexes in the sense of 7.2, while those motor skills which involve active movement rather than maintenance reflexes are generally learned in the manner described in 7.1.

7.3. *The cerebellar initiation of movements*

The two kinds of input–output relation give the cerebellum the power to learn any task whose execution is related in a rather rigid way to information sent through the mossy fibres, and at the same time to set up suitable reflexes to maintain balance and posture during execution of those tasks. The cerebrum is thus freed from at least the routine matters associated with motion and stance. There are, however, many instances in life when both the recognition that a job must be done, and its implementation, are simple operations. For example, information taken out of the visual system at a fairly low cortical level (say from areas 18 and 19) might be

useful as a source of cues during walking: and information about the mood
one is in can sometimes influence in a simple (but learned) way the
gestures one makes.

It is but a short step from believing that the cerebellum stores movements
and gestures to proposing that visual cues and information about mood and
so forth can form enough of a context actually to initiate an action; and it
would be strange if something of this sort did not happen, though it doubt-
less occurs more frequently in the motor cortex. Where it is possible to
translate the combined activity of many cerebral fibres rather simply into
physical directives, doing so in the cerebellum would free the cerebrum from
an essentially tedious task. In these circumstances, the cerebellum becomes
rather more than a slave which copies things originally organized by the
cerebrum: it becomes an organ in which the cerebrum can set up a sophisti-
cated and interpretive buffer language between itself and muscle. This can
be specially tailored to the precise needs of the animal, and during later
life leaves the cerebrum free to handle movements and situations in a
symbolic way without having continually to make the retranslation. The
automatic cerebellar translation into movements or gestures will reflect in a
concrete way what may in the cerebrum be diffuse and specifically un-
formulated, while the analysis leading to that diffuse and unformulated
state can proceed in its appropriate language.

§8. THE MAIN PREDICTIONS OF THE THEORY

8.1. *Modifiable synapses*

The main test of the theory is whether or not the synapses from parallel
fibres to Purkinje cells are facilitated by the conjunction of presynaptic
and climbing fibre (or post-synaptic) activity (5.1). If this is not true, the
theory collapses.

It is likely that no other cerebellar synapses are modifiable. The mossy
fibre-granule cell synapses are discussed in 3.3.3, and the Golgi cell afferent
synapses in 4.5. The function of the stellate cells is fixed throughout the
life of the cerebellum, and so they probably do not possess modifiable
synapses. Though it is difficult to see how these predictions could be wrong,
they might be: such a disproof would be embarrassing but not catastrophic,
since something of the bones of the theory would remain.

8.2. *Cells*

The roles of the various cells in the cortex are roughly determined once
the main prediction about modifiable synapses is verified. There are, how-
ever, three predictions which can be tested. The first concerns the Golgi
cells. They have been discussed at some length, and arguments were

produced for the view that there should be little if any summation between the upper and lower dendritic trees (4.4). The cell should be driven by that tree which is currently the more powerfully stimulated. Refutation of this would again be awkward but not fatal.

Secondly, the interpretation of the stellate cells as a threshold setting mechanism for the Purkinje cells depends strongly upon the presumed distribution of the mossy fibre rosettes below the cortex. The theory requires that each mossy fibre extends a fair distance perpendicular to the line of the folium (6.1), and this can be investigated.

Thirdly, the number of granule cells active at any one time (say in any 50 msec period) is a small fraction (less than 1/20) of all granule cells.

8.3. *Input–output relations*

The two forms of input–output relation are experimentally distinguishable, and the same olivary cell–Purkinje cell pair may at different times be used both ways. For the learned movement form, the olivary cell should respond to a command for the same elemental movement as is initiated by the corresponding Purkinje cell. For learned conditional reflexes, the activity provoked by the Purkinje cell must tend to cause a reduction in the receptor activity which drives the olivary cell.

It is proposed that these two input–output relations are used for fairly different tasks. This division of labour is not logically necessary, since in principle each form can execute either task: but it would be surprising if the observed division differed substantially from the one suggested, since that particular arrangement is the most economical.

§9. THE CODON REPRESENTATION

The notion central to the present theory is that the afferent input events communicated by the mossy fibres to cerebellar cortex are turned into a language of small subsets and then stored; and this has been called the codon representation of an input. This formulation is new, but the principle is closely related to the feature analysis ideas current in the machine intelligence literature (see e.g. Uhr & Vossler, 1961). 'Features' are merely rather specially chosen codons. This author in particular owes a debt to the paper by Brindley (1969) which contains what may be regarded as a degenerate case of codon representation, though from rather a special point of view.

The idea of codons arose in an unlikely way as the result of a search for a representation which the cerebrum might use for storing information. Its relevance to the cerebellum was noticed only when it was realized that any neural net built to implement the representation must contain something

like granule cells. An analysis of the properties of the codon representation and of its possible place in the theory of cerebral cortex will form the subject of a later paper.

REFERENCES

ARMSTRONG, D. M., ECCLES, J. C., HARVEY, R. J., MATTHEWS, P. B. C. (1968). Responses in the dorsal accessory olive of the cat to stimulation of hind limb afferents. *J. Physiol.* **194,** 125–145.

BRINDLEY, G. S. (1964). The use made by the cerebellum of the information that it receives from sense organs. *Int. Brain. Res. Org. Bulletin* **3,** 80.

BRINDLEY, G. S. (1969). Nerve net models of plausible size that will perform many of very many simple learning task. *Proc. R. Soc.* B. (In the Press.)

BRODAL, A. (1954). Afferent cerebellar connections. In *Aspects of Cerebellar Anatomy*, ed. JANSEN, J. & BRODAL, A. ch. II, pp. 82–188. Oslo: Johan Grundt Tanum Forlag.

CAJAL, R. Y. (1911). *Histologie du Système Nerveux*, Tome II, 1955 edn., p. 57, C.S.I.C.: Madrid.

ECCLES, J. C., ITO, M. & SZENTAGOTHAI, J. (1967). *The Cerebellum as a Neuronal Machine.* Berlin: Springer-Verlag.

ESCOBAR, A., SAMPEDRO, E. D. & DOW, R. S. (1968). Quantitative data on the inferior olivary nucleus in man, cat and vampire bat. *J. comp. Neurcl.* **132,** 397–403.

HEBB, D. O. (1949). *The Organization of Behaviour*, p. 62. New York: Wiley.

UHR, L. & VOSSLER, C. (1961). A pattern recognition program that generates, evaluates and adjusts its own operators. *Proc. west. jt. Computer Conf.* **19,** 555–569.

WALBERG, F. (1954). Descending connections to the inferior olive. In *Aspects of Cerebellar Anatomy*, ed. JANSEN, J. & BRODAL, A., ch. IV, pp. 249–263. Oslo: Johan Grundt Tanum Forlag.

ADDENDUM

Escobar, Sampedro & Dow (1968) have shown that in man, and probably also in cat, there are fewer cells in the inferior olive than there are cerebellar Purkinje cells. There may therefore exist other sources of climbing fibres. Statements in the present work about the inferior olive should be understood to refer to all sources of climbing fibres, including those as yet undiscovered. If olivo-cerebellar fibres are found to branch, the theory will require slight modification.

EXPLANATION OF PLATE

Dendritic spines on a cat Purkinje cell. (From Eccles *et al.* 1967, Fig. 27 A.)

Plate 1

10 μ

DAVID MARR