

PRIMARY AND SECONDARY WAVES IN
DEVELOPMENTAL BIOLOGY

By E.C. Zeeman

ABSTRACT.

Using catastrophe theory, we prove a theorem to the effect that whenever a multicellular mass of tissue differentiates into two types, the frontier between the two types always forms to one side of its final position, and then moves through the tissue before stabilising in its final position. We call this movement a primary wave. Primary waves may sometimes be identified as hidden waves of cell determination, which may not manifest themselves visibly until after a delay of several hours. The visible manifestation will then be a secondary wave of cellular activity, which may cause morphogenesis, for example rolling changes of curvature.

Two applications are worked out in detail, namely models for gastrulation and neurulation of amphibia, and for culmination of cellular slime mold. In the amphibian model the differentiation between ectoderm and mesoderm causes a hidden primary wave, whose visible secondary wave of cells submerging causes not only the morphogenesis of gastrulation but also the formation of notochord and somites. In the slime mold model the differentiation between spore and stalk causes a hidden primary wave, whose visible secondary wave of cells submerging causes culmination and the morphogenesis of the fruiting body. Both models suggest experiments by which they can be tested.

1.

INTRODUCTION.

Our objective is to explain primary waves by catastrophe theory [18], and secondary waves by cell physiology [1,8], and then to use them both together to explain morphogenesis.

By a wave we mean the movement of a frontier separating two regions. We call the wave primary if the mechanism causing the wave depends upon space and time. We call the wave secondary if it depends only upon time, in other words it is series of local events that occur at a fixed time delay after the passage of the primary wave. Therefore, whereas the wave-form of the primary wave is fundamental, the secondary wave only appears to have a wave-form because it follows the primary wave after the fixed time delay. In a sense the wave-form of the secondary wave is accidental because it could be disrupted by mixing up the substrate in between the passage of the two waves. The epidemic example in §2 below illustrates this point. The point is further emphasised by the following difference between the primary and secondary waves : if the substrate is cut before the passage of the primary wave then this stops the primary wave. However if the cut is made between the passage of the two waves then this will not stop the secondary wave, which will appear to jump across the cut.

If the primary wave is invisible, then the secondary wave may appear mysterious. We suggest that this may be a typical situation in developmental biology. For instance a primary wave across a multicellular mass of tissue might consist of the switching on of certain gene systems in each cell, and this may be difficult to detect at the time because biochemical analysis tends to disrupt the delicate dynamics; in

fact most experimental evidence that gene systems have been switched on seems to come from observation of some secondary effect after a suitable time delay (see §7 below). The secondary effect in this case is usually some physical manifestation in cell behaviour such as change in chemical composition, change in RNA content, change in oxygen consumption, change in membrane cohesiveness, change in shape, change in amoeboid activity, change in mitosis rhythm, etc. Another common and important secondary effect is for the cell to alter the ratio between the areas of that part of its membrane in contact with other cells and that part comprising free surface of the tissue; for instance the cell can increase contact with other cells by amoeboid action towards them, and decrease its free surface by wrinkling its free membrane (see Figure 21). For convenience we call this process submerging. For example submerging happens during gastrulation (see Figure 20).

It is the secondary wave of physical manifestation that may signal the release of chemical energy to provide the physical energy necessary for morphogenesis. For example submerging cells may push and pull on their neighbours, and thereby alter the overall curvature of the free surface, as described in Gustafson & Wolpert [8]. We suggest that some morphologies that hitherto may have appeared to be explicable may now be explained in terms of secondary waves. If this is the case, this may provide a conceptual framework for the experimental search for hidden primary waves.

The next question is : what causes a primary wave ? The simplest mechanism is diffusion, for instance of chemicals or signals. In order to illustrate the difference between primary and secondary waves, we briefly give elementary examples of epidemics and regulation in §§ 2 and 3, in which the primary wave is caused by diffusion.

However for the rest of the paper we are interested in a more subtle mechanism for producing primary waves. We prove a theorem that the four hypotheses

- I Homeostasis
- II Continuity
- III Differentiation
- IV Repeatability

together imply the existence of a primary wave. Another way of stating

the theorem is that whenever a frontier forms between two types of tissue, it first forms off to one side, and then moves as a primary wave through the tissue before stabilising in its final position. The theorem gives no indication of the extent of the movement, although in some applications it appears the movement can be very large, over half the diameter of the embryo. Since the formation of frontiers is common in developmental biology, we should expect such primary waves to be common, and hence morphogenesis to be commonly caused by their secondary waves.

We state the theorem more precisely in §5, and prove it in §8. The proof consists of translating the above four hypotheses into mathematics, thereby making them precise, and then using Thom's classification of elementary catastrophes [18]. Since the latter result is deep, the theorem is a non-trivial description of the mechanism underlying this type of primary wave.

One of the interesting features of the theorem is that diffusion may or may not be present during the passage of the wave : it is irrelevant to the proof. In other words the complexity of biochemical events associated with a particular primary wave, even one that subsequently slows down to halt, may include, for instance, diffusion of chemicals across cell membranes, or diffusion of dynamical signals across the cells entraining some activity in them. On the other hand another primary wave may progress without any diffusion, and without any signals. In this case the wave would be purely kinematic, with each cell behaving according to its own internal clock. The clocks may have been synchronised initially, but, due to some underlying gradient across the tissue, may tick at different speeds. Therefore a switch inside a cell, for instance the hidden switching on of some gene system (analogous to the clock striking), will occur in different cells at different times. The continuity of the underlying gradient will ensure a continuity of these different times across the tissue, and so the switch progresses as a primary wave. The question remains as to what actually causes the switch, and an analysis of the proof of the theorem reveals that the basic cause is, surprisingly, homeostasis. What is homeostasis? - we choose to translate homeostasis into mathematics as a stable equilibrium point of a time-dependent multidimensional dynamical system (see §8).

Returning to diffusion for a moment : of course the continuity of the underlying gradient may itself be due to a much earlier diffusion, for instance in the cytoplasm of the original egg before cleavage. And again if a cell is grafted to alien surroundings, then a new diffusion may occur that upsets its clock. However under normal conditions, it could be possible for both development and primary waves to occur without diffusion. Therefore one implication of the theorem is to alter the possible expectations of the experimentalist concerning "morphogens". A related result from catastrophe theory [26] guarantees the existence of morphogens, but they may not act like classical "organisers". In other words, if we locate the organising centre of some morphology, we may still expect to find a morphogen, that is a chemical, physical or dynamic gradient, whose discontinuities reflect the organisation, but we should not necessarily expect to find an organiser, that is a chemical, physical or dynamic signal emanating from the centre.

The theorem gives both qualitative and quantitative predictions about the shape and speed of primary waves, and consequently also about their secondary waves. We illustrate the theorem by two applications in §§10 - 17, namely the gastrulation of amphibia, and the culmination of slime mold. In both examples we take as hypothesis (for which there is experimental evidence) a secondary wave of cells submerging. In the first example the differentiation between ectoderm and mesoderm causes a primary wave, by the theorem. We take as hypothesis (for which there is experimental evidence) that this hidden primary wave begins at the bottom of the grey crescent. We then deduce that the secondary wave causes, or helps to cause, gastrulation, the dorsal lip, the blastopore, the archenteron roof, the separation of mesoderm from endoderm, the neural folds, and the formation of notochord and somites during neurulation. The fact that nearly all the main morphogenetical movements arise from a single secondary wave, resulting from a single differentiation and the local activity of cells submerging, help to explain why these movements are common to many species.

In the second example of slime mold the differentiation between spore and stalk cells causes a primary wave, by the theorem. We take as hypothesis (for which there is experimental evidence) that the primary wave begins at the tip and proceeds $\frac{1}{3}$ of the way along the grex, several

hours before culmination. We then deduce the morphogenesis of the fruiting body, giving predictions as to shape and speed.

I am indebted to many people for discussions, particularly mathematicians, René Thom, David Fowler and Klaus Jänich, and biologists C.H. Waddington, Jack Cohen, Lewis Wolpert, Peter Nieuwkoop, Jonathan Cooke and John Ashworth. The main inspiration came from years of conversations with René Thom about applying catastrophe theory to biology. Meanwhile in counterconversations Lewis Wolpert emphasised the inadequacy of using catastrophe theory by itself, because it can only explain the geometry, and not the forces that shape the embryo. On the other hand looking at the local forces by themselves cannot explain the global geometry. Hence the concept of primary and secondary waves grew out of trying to put these two ideas together, the mathematical and the biological.

Discussions with Peter Nieuwkoop about gastrulation were particularly valuable during the germination of the ideas, after an initial presentation of the theorem at a conference in Göttingen in September 1973 organised by Klaus Jänich. Jonathan Cooke stimulated the ideas about pattern-formation and the somites. John Ashworth explained to me the slime mold morphogenesis. I am indebted to the A.A.A.S. and A.M.S. for the opportunity to present the ideas, and to the various authors and journals for permission to reprint their diagrams.

CONTENTS.

1. Introduction.
2. Example : epidemic.
3. Examples : regulation.
4. Example : ecology.
5. Primary waves in embryology.
6. Secondary waves in embryology.
7. Experimental detection of a primary wave.
8. Proof of the theorem.
9. Quantitative aspects of the theorem.
10. Gastrulation of amphibia.
11. Curvature.
12. Neurulation.
13. Pattern formation.
14. Timing of gastrulation and neurulation.
15. Experiments to be done.
16. Slime mold culmination.
17. Experiments to be done.
18. Conclusion.
19. References.

2. EXAMPLE : EPIDEMIC.

This is a simple example to illustrate the difference between primary and secondary waves. The substrate is people. The two regions are those who have been infected by the epidemic and those who have not. The frontier bounds the infected region. This frontier moves forward as a hidden primary wave of infection. This is a simple diffusion wave, that moves steadily forward at the speed of diffusion as each person infects his neighbours (similar to Huyghens' principle). Then after a fixed time delay the visible secondary wave of symptoms follows, assuming that the people have remained stationary. If the people move about in between the waves then the wave-form of the secondary wave will be disrupted. If the substrate is cut before the arrival of the primary wave, in other words the infected population is quarantined, then this stops the primary wave, the spread of infection. However if the substrate is cut between the waves, in other words only the population already showing symptoms is quarantined, then this does not stop the secondary wave, the spread of symptoms.

3. EXAMPLES : REGULATION.

In the heartbeat the pacemaker wave causing muscular contraction is a primary wave, and about half a second later the wave muscular relaxation is a secondary wave [24]. In the nerve impulse the membrane depolarisation along an axon is a primary wave, and about a millisecond later the repolarisation is a secondary wave. On both these cases the primary wave is electro-chemical, and could be described as a diffusion wave (diffusion of electrons) proceeding according to Huyghens' principle.

The main feature of a diffusion wave is that it proceeds at constant speed. By contrast the main feature of the more subtle kind of primary wave that the theorem describes, and which we shall be considering from now on, is that it slows to a halt; in other words the frontier stabilises.

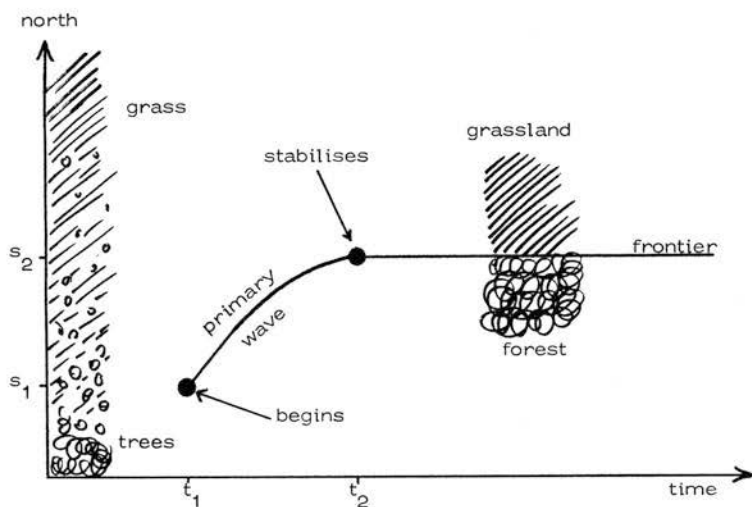
4.

EXAMPLE : ECOLOGY.

This is a simple example to illustrate the more subtle kind of primary wave. The wave has complicated causes, but is easy to understand because it is visible rather than hidden. The ecology is greatly oversimplified, but then we are only using it to illustrate the idea.

Consider the ecological development of grass and trees over a continuous environment of soil and climate. Suppose for simplicity that the northern end of the environment is suitable for grass only and the southern end for trees only, so that the former eventually develops into mature grassland, and the latter into mature forest. Suppose that as either vegetation gets established it suppresses the other; trees fail to survive in grassland and grass fails to survive in forest. Suppose at first there is a continuous variation of vegetation, varying from forest in the south, with trees gradually thinning as we proceed north, until grassland is reached. Then at time t_1 the forest will develop a noticeable frontier at latitude s_1 , say. This frontier will deepen, in the sense that the difference between the two sides of the frontier will become more marked, due to the suppressive effect of either vegetation upon the other. As the frontier deepens it would be exceptional for it to remain at s_1 , the place where it originally formed (exceptional from the point of view of repeatability, as we explain below). Therefore, depending upon the initial conditions, it will either move north as the mass of trees seed themselves into the grassland, or move south as the grassland erodes the forest edge. Suppose that in our case the initial conditions are such that the frontier moves north as in Figure 1.

Figure 1. The frontier of a forest moving as a primary wave.



Eventually the northerly expansion of the forest balances out against the unsuitability of the northern climate for trees, and so the northerly movement of the frontier slows down, until it stabilises at s_2 at time t_2 . Thereafter the frontier remains in stable equilibrium, and deepens further. The movement of the frontier from s_1 to s_2 during the time interval $t_1 < t < t_2$ is the primary wave.

The primary wave is succeeded by a series of secondary waves representing the spread of various species of flora and fauna that require various time delays of maturity before the forest becomes a suitable habitat for them (woodworm prefer old trees). However in this example we are less interested in the secondary waves.

To show that the primary wave is an illustration of the theorem, we must interpret the four hypotheses in this case.

I : Homeostasis is the tendency of the vegetation in any one place to develop into a stable state, stable with respect to time, that we can name as grassland or forest.

II : Continuity refers to both the continuous environment of soil and climate, and the initial continuous variation of vegetation.

III : Differentiation means that at the end there are two distinct states, mature grassland and mature forest.

IV : Repeatability means that if the initial conditions are varied slightly then the values s_1 , t_1 , s_2 , t_2 may vary slightly but the qualitative behaviour of the frontier remains the same. In other words repeatability means that the whole space-time development, or chreod [22], will be stable under sufficiently small perturbations of the initial conditions. Repeatability is an essential hypothesis for the existence of the wave, because it implies $s_1 \neq s_2$. Otherwise, if $s_1 = s_2$, meaning that the frontier had stabilised where it formed, and so causing no wave, then this would be unrepeatable, in the sense that the initial conditions must have been exceptional, and an arbitrarily small perturbation of them could cause $s_1 \neq s_2$, and hence cause a wave, in other words a qualitatively different development.

Notice that in this example of a primary wave, the wave could be said to be caused by diffusion as the trees seed themselves into the grassland. However the situation is not as simple as in the previous examples, where the speed of the wave was constant and equal to the speed of diffusion, because here the wave slows down and stops as the frontier stabilises. One could make an elementary model of this slowing down by using a linear differential equation with a diffusion term balanced against a survival term, but this would not give insight into the formation and the deepening of the frontier, as does the more sophisticated catastrophe model. Also the two models give different quantitative predictions, which would distinguish between them : for instance in the elementary diffusion model the wave slows down exponentially, but in the catastrophe model it slows down parabolically (see §9). The catastrophe model is more likely to be correct, because of the two processes involved, the initial seeding by diffusion, and the

eventual development into mature forest by homeostasis, the latter is the more significant.

To illustrate examples of kinematic waves that depend only upon gradients and internal clocks and not upon diffusion, consider the effects of latitude. For instance the spring blossoming of trees is a wave moving north (in the northern hemisphere) and in autumn the onset of migration by birds is wave moving south.

The reader will easily recognise the existence of many visible primary waves of this nature in ecology, evolution, anthropology and sociology. However in this paper we are more concerned with hidden primary waves in developmental biology, that cause secondary waves of physical manifestations in cells, that in turn may cause morphogenesis. We shall therefore state and prove the theorem in this context.

5. PRIMARY WAVES IN EMBRYOLOGY.

We begin by enlarging a little on the meaning of the four hypotheses in developmental biology. There is no need to give precise definitions at this stage, because the terms are given precision by the way we choose to translate them into mathematics in the proof of the theorem in §8 below. Suppose that E is a multicellular mass of tissue. We are concerned with development of E during a particular time interval T .

I : Homeostasis means that each cell is in stable biochemical equilibrium, an equilibrium that may change with time.

II : Continuity means that at the beginning of T we can represent the chemical, physical and dynamical conditions in different cells by smooth functions on E (the conditions inside a particular cell are represented by the values of the functions at the centre of mass of that cell). In an embryo, where the tissue has developed from an egg by cleavage, the continuity is inherited from the original continuity in the egg, which was due to diffusion in the egg cytoplasm. Any slight discontinuities that arise later tend to be evened out by subsequent diffusion across the cell walls. In aggregates of cells like slime mold, continuity means that the cells have sorted themselves out according to continuous gradients during the aggregation process.

Continuity implies that neighbouring cells will follow nearby paths of development whenever possible. We shall prove that where a

frontier stabilises this is not possible, and so across the frontier neighbouring cells will follow divergent paths of development, and large discontinuities will therefore arise.

III: Differentiation means that, whereas at the beginning of T there is only one type of cell (or, more precisely, a continuous variation amongst the cells), at the end of T there are two distinct types, and no continuous variation from one type to the other.

For simplicity we may assume that the tissue E is polarised, that is to say all variation takes place in one direction only (like the north-south line in the previous example in §4). Therefore for mathematical analysis it suffices to consider a 1-dimensional space interval S in that direction. Continuity means that at the beginning of T the cells vary continuously along S . Differentiation means that during T the cells at opposite ends of S develop continuously into different types. At the end of T , since there is no continuous variation between the two types there must be a frontier point in S separating the two types. This implies a frontier surface in E , separating the two types of tissue. If we can show that the frontier point in S moves, then this will imply that the frontier surface in E moves.

IV: Repeatability means that the development is stable, that is to say a qualitatively similar development will take place under sufficiently small perturbations of the initial conditions.

Main Theorem. Homeostasis, continuity, differentiation and repeatability imply the existence of a primary wave. In other words a frontier forms, moves and deepens, then slows up and stabilises, and finally deepens further.

Therefore whenever a frontier forms, it first forms off to one side and then moves as a primary wave through the tissue before stabilising in its final position. Here by "final position" we mean the position relative to the underlying tissue, which itself may be undergoing morphogenetical movements. The theorem is illustrated in Figure 1, and the proof is given in §8 below.

Remark 1. The theorem is qualitative rather than quantitative; in other words it is a result invariant under diffeomorphisms of space and time. Therefore the theorem cannot predict the extent of travel of the wave, and so in applications the extent must always be taken as an extra

hypothesis, and verified experimentally as in §7 below. It appears that some primary waves may travel a large distance, particularly those associated with morphogenesis. For example in gastrulation of some newts the ectoderm/mesoderm frontier travels from the grey crescent at latitude 40°S (see Figure 12) to its stabilisation position at 40°N , which is more than half the diameter of the blastula.

On the other hand the theorem does give quantitative predictions about the initial deepening, and the final stabilisation, of the frontier, because both these obey parabolic laws, and parabolicity is a diffeomorphism-invariant (see §9). These laws should furnish easily testable predictions.

Remark 2. The theorem gives no indication of whether the primary wave is visible or hidden. In embryology primary waves are generally hidden, in the sense of being experimentally undetectable at the time, because they probably consist of the switching on of gene systems, although their passage can sometimes be tracked in retrospect by the grafting experiment described in §7 below.

6. SECONDARY WAVES IN EMBRYOLOGY.

The theorem gives no indication of whether or not a hidden primary wave will result in a visible secondary wave after a time delay; and if it does, the theorem gives no indication of the size of delay, nor of the type of secondary wave. These must depend upon extra detailed biochemical hypotheses about that particular systems that are switched on, what the long term effects these systems have on the cells, how these effects physically manifest themselves, and whether there is a resulting energy release. Broadly speaking these are three possibilities.

In order to describe the three cases it is necessary to be a little more precise about what we mean by the word differentiation. Differentiation can be used in two senses, firstly the hidden determination of the cells into two types that takes place during the passage of the hidden primary wave, and secondly the subsequent development of physical difference between the two types that can be observed. In the hypothesis of the theorem in the last section we used differentiation in

the first sense, because this was the best word to capture hypothesis III, the determination of distinct types. In this and the next section we use differentiation in the second sense, because this is the more normal usage. We now describe the three cases.

(i) There is no secondary wave of energy release, and the speed of differentiation is slow compared with the speed of the wave. No release of energy implies no morphogenesis. The slowness of differentiation implies that any secondary wave will probably be unnoticeable. Therefore the only visible effect will be the slow appearance of the frontier between the two types of tissue in its final position s_2 . The original primary wave may never be noticed unless looked for.

(ii) There is no secondary wave of energy release, and the speed of the differentiation is fast compared with the speed of the wave. Again no release of energy implies no morphogenesis. However in this case the swiftness of differentiation implies that the frontier between the two tissues will appear before it has stopped moving, and so will present a visible secondary wave. The frontier may not necessarily first appear at s_1 , because it may not yet be deep enough to notice, but it will appear at some point s_3 , where $s_1 < s_3 < s_2$. It will then move towards s_2 , deepening and slowing down according to a parabolic law. This means that near s_2 the speed of the wave s is proportional to $\sqrt{s_2 - s}$ (see §9 Corollary 3). The frontier then stabilises at s_2 and deepens further.

This is a commonly observed phenomenon, and is often described as recruitment. For example* a developing insect eye starts with a few cells, and then enlarges by recruiting neighbouring epidermis cells. The word "recruitment" implicitly suggests that the experimenter should look for an "evocator" or an "organiser" emanating from the existing eye cells causing the recruitment. However if this expanding frontier of the eye were a secondary wave, then perhaps one ought to look for a hidden primary wave passing some hours before, without necessarily any organiser. In mammalian eyes the wave goes the other way: instead of expanding outwards the optic region shrinks in size.

* I am indebted to Peter Shelton for this example.

(iii) There is secondary wave of energy release. In this case there will be a dramatically observable secondary wave causing morphogenesis. Geometrically the situation will be complicated by the fact that not only is the wave moving through the tissue, but the tissue itself is also moving and changing shape, with the former causing the latter. It is also likely that the energy release precedes differentiation. Therefore differentiation may (or may not, as in cases (i) and (ii) above) appear as another secondary wave after the morphogenesis. Therefore two misinterpretations are possible: Firstly, it may be geometrically unobvious to relate the visible secondary differentiation wave to the hidden primary wave, because the morphogenesis will have moved the tissue around in between the two. Secondly it may be tempting to conclude that the differentiation has been evoked by the morphogenesis, or by the new position of the tissue, rather than by the original hidden primary wave. Thus the experimenter's path may be strewn with pitfalls, unless he manages to uncover the primary wave, which possibility we now discuss.

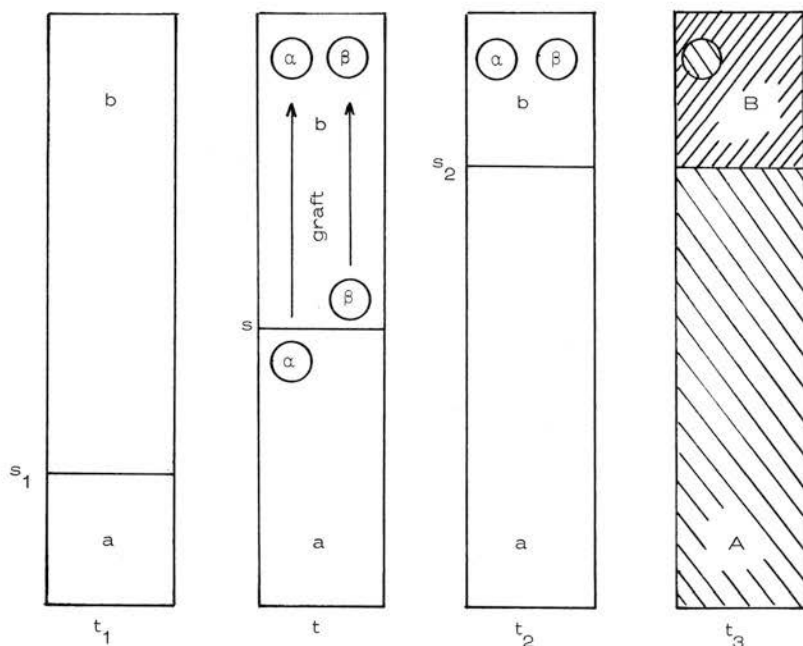
7. EXPERIMENTAL DETECTION OF A PRIMARY WAVE.

Sometimes the hidden **primary** wave may mark a loss in potentiality. In this case, the passage of the wave can be detected by a standard grafting experiment, provided that the tissue is suitable for grafting. For example one can detect the hidden mesoderm wave in amphibian gastrulation by this method*.

Suppose, as before, that the wave starts at s_1 at time t_1 and stabilises at s_2 at time t_2 . At this stage it is hidden and so there is no detectable difference between the cells, but eventually at some later time t_3 differentiation will cause a physically observable difference that we indicate by shading the regions marked A, B in Figure 2. At time t_1 , although there is no difference, we can say that a-cells are presumptive A-cells and b-cells are presumptive B-cells. The wave travels from a towards b. The passage of the wave past a cell is indicated by that cell switching from being a b-cell to being an a-cell.

*I am indebted to P.D. Nieuwkoop [12] for explaining this experiment to me.

Figure 2. Detecting a primary wave by grafting.



The switch occurs essentially because of homeostasis, as we explain in the proof of the theorem in the next section. Therefore b-cells have the potentiality to develop in either A-cells or B-cells. However a-cells may only have the potentiality to develop into A-cells, and we suppose that this is the case. Therefore the primary wave marks the loss in B-potentiality.

If we want to verify that the hidden primary wave has reached position s at time t , then at time t graft two small pieces α , β from just behind and just ahead of s onto b-tissue, well clear of the wave, as shown in Figure 2. The β -cells are influenced by their new position to remain b-cells, and so by time t_3 develop into B-cells, causing the β -graft to disappear. Meanwhile the α -cells have already switched and lost their B-potentiality and so by time t_3 the α -graft stand out as a patch of A against B.

8. PROOF OF THE THEOREM.

We follow the conceptual ideas of Thom [18, 19]. As explained in §5 above it suffices to use a 1-dimensional interval of space S , transverse to the forming frontier. Let T be a time interval encompassing the development. Let $C = S \times T$ be the rectangle of space-time. Let X denote a manifold representing the states of a cell. One can envisage X as a bounded open subset of n -dimensional euclidean space R^n , where n may be very large (possibly several thousand). The coordinates $\{x_i; i=1,2,\dots,n\}$ of a point $x \in X$ may represent not only the concentrations of the different proteins in the cell, and the rates of change of those concentrations, but also may include variables representing various physical characteristics of the cell, its membrane, the cell dynamics, etc.

Consider a cell at the point $c \in C$. By Hypothesis I this cell is in homeostasis. We choose to translate homeostasis into mathematics by assuming that the biochemistry of the cell can be modelled by a gradient dynamical system on X

$$\dot{x} = - \text{grad } V_c,$$

where $V_c : X \rightarrow R$ is a smooth function and R denotes the real numbers (see Remark 3 below for the meaning of this function). We choose to translate Hypothesis II, continuity, into mathematics by assuming that V_c can be chosen to depend smoothly on c . Therefore we have a function

$$V : C \times X \rightarrow R$$

given by $V(c,x) = V_c(x)$. We choose to translate Hypothesis IV, repeatability, into mathematics by assuming that V is generic*. Let $M \subset C \times X$ denote the set of stationary values of V , given by $\nabla V=0$, where ∇ denotes the gradient with respect to X . Let G denote the closure of the subset of minima, which are given by $\nabla^2 V$ positive definite, where ∇^2 denotes the Hessian with respect to X . Then by smooth genericity, M is a smooth 2-dimensional surface in the $(n+2)$ -dimensional space $C \times X$, and G is a subsurface of M with boundary ∂G . Let $\chi : M \rightarrow C$ denote the

* Generic means in general position, that is to say the map $c \rightarrow V_c$ maps C transverse to the natural stratification of $C^\infty(X)$. Generic V 's are open dense in the space of all V 's, and therefore both stable, and permissible to use as models.

map induced by the projection $C \times X \rightarrow C$. By Thom's classification theorem of elementary catastrophes [18], the only singularities of χ are fold curves and cusp points, since V is smooth and generic. The boundary ∂G consists of fold curves and cusp points.

By homeostasis the state of cell at c is at a minimum of V_c , and therefore represented by a point

$$\sigma(c) \in G \cap \chi^{-1}c.$$

Therefore σ is a section of χ ,

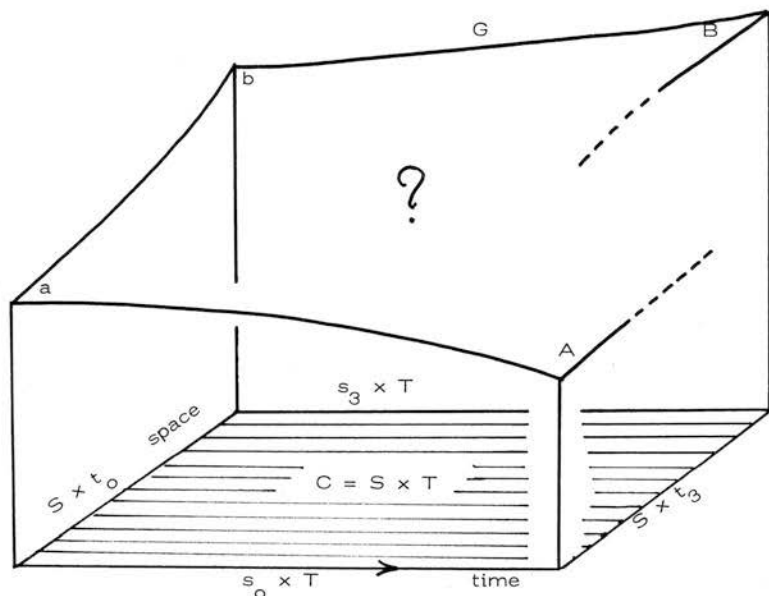
$$\chi \begin{array}{c} \downarrow \\ \sigma \end{array}$$

In other words $\chi\sigma = 1$. The interesting point is that, whereas χ is smooth (induced by projection), σ may be forced by χ to be discontinuous. We now use Hypothesis III, differentiation, to analyse this discontinuity and show, using I and IV again, that it implies the primary wave.

Let $S = [s_0, s_3]$, $T = [t_0, t_3]$. Let us analyse the continuity of σ on the boundary ∂C of $C = S \times T$. Firstly σ is continuous along the side $s_0 \times T$ because a-cells are developing smoothly into A-cells. Similarly σ is continuous along $s_3 \times T$ because b-cells are developing smoothly into B-cells. Next σ is continuous along the side $S \times t_0$, because, by Hypothesis II, continuity, we may assume that at the beginning the tissue is continuous.

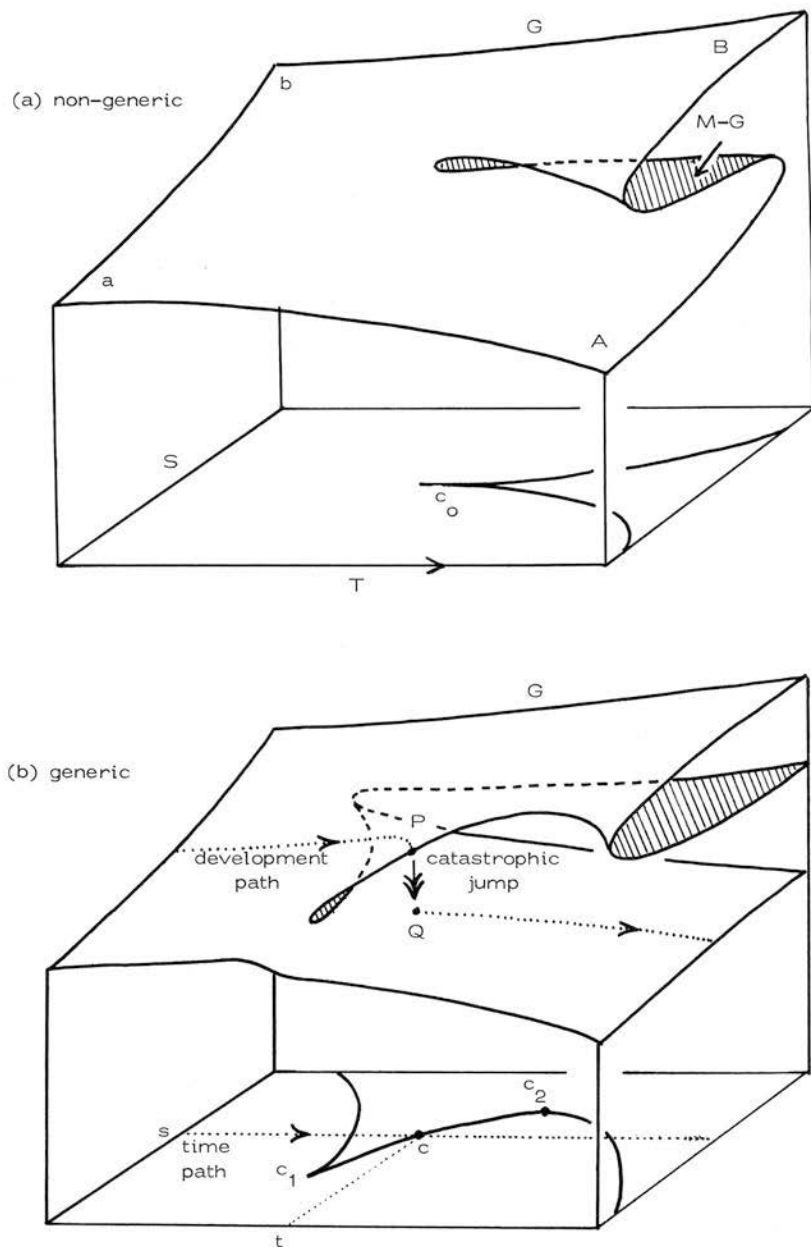
Finally σ cannot be continuous along the side $S \times t_3$, because, by Hypothesis III, differentiation, at the finish two distinct types of cells A, B have developed, with no continuous variation between them. By Hypothesis II, continuity, A must spread continuously from one end, and B from the other, towards some point of discontinuity, which is, as yet, undetermined.

Figure 3 shows the partial section of G over ∂C that must exist ready to receive the map $\sigma : \partial C \rightarrow G$. Now comes the problem of extending σ to the interior of C . First we ask the simpler question: what singularity must the map $\chi : M \rightarrow C$ have over the interior of C ? Since $M \supset G$, M must extend the section over ∂C shown in Figure 3. Therefore by the classification theorem M must have at least one cusp singularity over the interior of C . A single cusp would be sufficient. Moreover a single cusp is qualitatively the simplest solution of the

Figure 3. Section of G over the boundary of $S \times T$.

extension problem, and we can justify the simplest solution by again appealing to Hypothesis II, continuity; in other words we assume minimal discontinuity subject to Hypothesis III, differentiation.

In Figure 4 we illustrate two examples of a surface M over C , each with one cusp, and each extending the given section over ∂C . In each case the shaded subsurface indicates $M-G$ (representing saddle-points of V , in other words unstable equilibria of the biochemical dynamic, and so not realisable by homeostasis). If we ignore the product structure of space-time $C = S \times T$, then the two pictures are qualitatively equivalent. However if we take note of the product structure, then Figure 4a is exceptional because the time-axis at the cusp point c_0 coincides with the cusp-axis. Hypothesis IV, repeatability forbids this exceptional

Figure 4. The graph G of homeostatic states.

situation and so Figure 4a is ruled out.

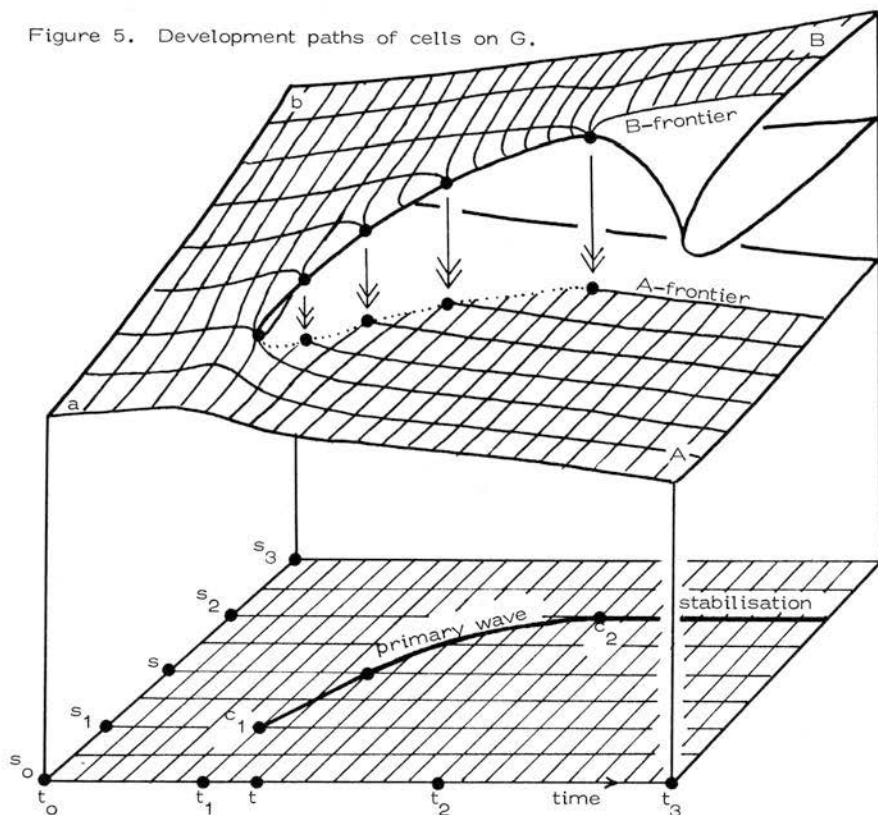
The reader may ask why do we choose the rather elaborate Figure 4b, and so we must explain. Firstly the direction of the cusp-axis must have a non-zero S-component (to avoid the fault of Figure 4a) and a positive T-component, because the frontier between the differentiated tissue must occur inside the cusp. Probably in most cases the T-component will be greater, but to emphasise both components in Figure 4b we have drawn them approximately equal, in other words in the perspective drawing of C we have drawn the cusp-axis inclined at 45° to the time-axis at c_1 . Secondly, as the two branches of the cusp widen out, there will be a unique first point c_2 (on the branch for which t is greater) where the tangent is parallel to the time-axis. These two points c_1, c_2 will mark the beginning and ending of the primary wave, as we shall now prove. We have merely drawn Figure 4b so as to emphasise the qualitative features of these two points.

We now plot the development of each cell by lifting its time-path in C up onto the graph G. Since G is the graph of homeostatic states, the lifted path will represent how the state of the cell changes, in other words will represent its development-path (see Figure 4b).

Since homeostasis is represented by a differential equation $\dot{x} = -\nabla V$, the development path will be a continuous path on G, held continuously in stable equilibrium by the differential equation, unless the path happens to cross ∂G . Therefore, in the language of Thom [18], the changing state will obey the Delay Rule. Now ∂G is the fold curve of M lying above the two branches of the cusp. Suppose that the time-path of the cell at s crosses the branch c_1, c_2 of the cusp at the point c at time t (see Figure 4b). Then the development-path of s will cross ∂G at the point P above c, and at this point the homeostatic stability breaks down, because the corresponding minimum of V_c has coalesced with a saddle (represented by M-G), and disappeared. Consequently the homeostatic differential equation comes into play and carries P rapidly to Q, which is the unique new stable equilibrium on G above c, in whose basin of attraction P lies (see [26]). The rapid change of state from P to Q caused by homeostasis is called a catastrophe, or catastrophic jump; this is the moment when the b-cell at s switches into an a-cell. What we have previously vaguely referred to as "switching on of gene systems" is

represented mathematically by the fast flow from P to Q along an orbit of the differential equation in X representing homeostasis. That was why in the introduction and in §7 above we remarked that the switch occurs essentially because of homeostasis. Therefore at time t the cell s marks the frontier between a-cells and b-cells, namely the position of the wave. The depth of the frontier is represented by the length PQ . The catastrophe occurs in all cells lying in the interval $s_1 < s < s_2$ during the period $t_1 < t < t_2$, as their time-paths cross the cusp branch $c_1 c_2$, and this defines the primary wave. Therefore $c_1 c_2$ determines the track of the primary wave in space-time. Meanwhile cells for which $s < s_1$ suffer no catastrophe, but develop from a-cells into A-cells along a smooth development-path. Similarly cells for which $s > s_2$ develop smoothly from b-cells into B-cells. The various development paths are shown in Figure 5.

Figure 5. Development paths of cells on G .



From Figure 5 we can read off the qualitative features required in the statement of the theorem, as follows :

At time t_1 the frontier first forms at s_1 .

Between t_1 and t_2 the frontier moves from s_1 to s_2 and deepens.

At the time approaches t_2 the frontier slows up, and approaches s_2 .

At t_2 the frontier reaches s_2 , and stabilises.

After t_2 the frontier deepens further.

This completes the proof of the theorem. Before we proceed to quantitative features of the theorem we make four remarks.

Remark 1. Sometimes the wave does not begin in the middle of the tissue, but on the boundary of the tissue, so that the tissue appears to "grow into" the frontier. This seems to be the case with slime mold (see §15 below), and with the development of chicken wings, for instance. In this case the mathematics is simpler because the space-time track of the tissue does not cross the cusp point, but only the fold curve. In fact there may not necessarily be any cusp point at all. In Figure 5 the slime-mold grex would be represented by $[s, s_3]$ with tip at s and tail at s_3 . At time t the wave would begin at the tip, and then proceed along the grex to stabilise at s_2 at time t_2 . The front part $[s, s_2]$ eventually develops into stalk-cells A, and the back part $[s_2, s_3]$ into spore-cells B.

Remark 2. Not all frontiers are formed by primary waves of this type, because in some cases our translation of the four hypotheses into mathematics may not be valid. For example in the gastrulation of birds and mammals, or in mixing experiments [1,10], the frontier is caused by migration of different types of cells, sorting themselves out, whereas we have assumed that the cells stay more or less in the same place relative to one another in the tissue. However in some of these cases a primary wave may already have taken place in some underlying gradient, and the migration of cells up or down the gradient may be merely a secondary wave.

Remark 3. There was one drastic simplification that we made in the proof of the theorem, in the way that we chose to translate homeostasis into mathematics. It may well be reasonable to represent homeostasis by a dynamic D on X , but it is not obvious that D should be a gradient dynamic, $\dot{x} = -\nabla V$. In special cases V may represent some

potential energy that is minimised, and then it is reasonable. But in general D may be non-gradient, particularly when the cell contains biological clocks. Even then in some cases it is technically possible to reduce D to the gradient case, by choosing V to be a Lyapunov function for D (see [18, 26]). However in other cases this may not be possible; frontiers arising from, or associated with turbulence, for instance, would probably not behave so simply.

Remark 4. We have drawn Figures 3, 4, 5 as if X were 1-dimensional, and as if

$$M \subset C \times R \subset R^3.$$

In fact this is not true because X is an open subset of R^n , where n may be very large, and therefore more precisely

$$M \subset C \times X \subset R^{2+n}.$$

However this does not alter the fact that M is 2-dimensional surface, and therefore our diagrams are indeed rigorous pictures of the map $\chi : M \rightarrow C$. Moreover Thom's classification theorem [18] can be modified [26] in an important manner for this context, as follows :

If $\chi : M \rightarrow C$ has a cusp catastrophe, then in the neighbourhood of that point we can choose a map $\pi : X \rightarrow R$ such that

$$1 \times \pi : C \times X \rightarrow C \times R$$

throws M diffeomorphically onto the surface pictured in Figure 4b. Moreover π can be chosen to be the projection of X onto one of the given axes of R^n , that measure concentrations etc. In fact we can choose any axis not perpendicular to the tangent to ∂G at σ_1 . Let us call the chemical or physical property that this axis measures, a morphogen. Then the vertical axis in Figures 3, 4, 5 measures the morphogen. The morphogen need not be unique, and may only be an artifact. But if the morphogen is easy to measure, then it may be useful for experimental predictions. It is remarkable that Thom's theorem guarantees the existence of a morphogen for each developing frontier.

9. QUANTITATIVE ASPECTS OF THE THEOREM.

From Figure 5 we can deduce some quantitative estimates about primary waves. The estimates are computed to first order in small quantities, and are therefore only accurate near the beginning and

the end of the wave.

Corollary 1. Initially, when the frontier first forms, it is moving at constant speed.

Proof. The path of the wave is the branch c_1c_2 of the cusp, which, near c_1 , to first order, can be replaced by the tangent at c_1 .

Corollary 2. Initially, when the frontier first forms, its depth increases by a square-root law, in other words the depth of the frontier is proportional to $\sqrt{t-t_1}$ (and hence also to $\sqrt{s-s_1}$).

Proof. At time t the depth of the frontier is equal to the catastrophic jump PQ in Figure 4. Therefore we must compute PQ . Choose origin O at sc_1 , the point of M over the cusp point c_1 . Choose two axes ξ, η at O as follows: ξ is measured along the tangent at O to ∂G in X , oriented towards P , and η is measured along the tangent at c_1 to the cusp in C . Let K denote the (ξ, η) -plane. Then K is the osculating plane of ∂G at O . Therefore, by genericity, and ignoring third order terms, ∂G lies in K and has equation $\eta = k\xi^2$, where $k > 0$. Therefore P satisfies $\xi = +\sqrt{\eta/k}$. Meanwhile Q satisfies $\xi = -2\sqrt{\eta/k}$, because M is the diffeomorphic image of a cubic surface, namely the canonical cusp catastrophe. Therefore $PQ = 3\sqrt{\eta/k}$. But η is proportional to $t-t_1$, and hence PQ is proportional to $\sqrt{t-t_1}$, as required.

Remark. The initial position and movement of a developmental wave may be difficult to observe, because of the initial shallowness of the frontier. However it might be possible to find the initial position by using Corollary 2 to extrapolate backwards (and hence find the organising centre, if the wave happens to emanate from a point).

Note that Corollary 1 also remains true for any secondary wave. However we should not necessarily expect Corollary 2 to apply to a secondary wave, because the two waves are of a totally different nature: the primary wave marks the frontier between two diverging types of tissue, whereas the secondary wave makes the onset of a secondary effect within one type of tissue. Therefore the initial movement of the primary wave may be difficult to observe, whereas that of the secondary wave may be easy to observe - for instance the first invagination in gastrulation (see §10 below).

Corollary 3. Eventually, just before the frontier stabilises, it slows down parabolically. In other words $(s_2 - s)$ is proportional to $(t_2 - t)^2$, and the speed is proportional to $(t_2 - t)$.

Proof. By genericity the curve $c_1 c_2$ touches the time axis at c_2 with quadratic tangency. Therefore near c_2 , ignoring third order terms, the curve has the equation

$$(s_2 - s) = h(t_2 - t)^2,$$

where $h > 0$. The speed is given by differentiating :

$$\dot{s} = 2h(t_2 - t) .$$

This completes the proof of Corollary 3.

Remark. Corollary 3 should be easy to observe and verify. If there is no morphogenesis, then the same result will hold for any secondary wave. Therefore in a recruitment phenomenon, for instance, the parabolic law might provide a good test to distinguish whether it was a secondary wave or the result of entrainment.

If there is morphogenesis, then the displacement of cells relative to one another may upset the parabolic law for the secondary wave, but Corollary 3 may nevertheless yield other quantitative predictions - see for example the estimate for stalk diameter in the slime mold fruiting body, in §16 below.

Energy release. Some morphogenetical movements begin slowly, and build up to a recognisable climax before finally dying down. This can be seen most clearly in time lapse films, and the reader is especially recommended to see the two Göttingen films of Luther [11] on gastrulation and neurulation, and Gerisch [7] on slime mold. For instance in [11] gastrulation begins by invaginating slowly, then the tissue begins to roll over the dorsal lip, then pours over the entire circle of blastopore lip, until it eventually slows down, and the blastopore gently closes. Similarly neurulation begins with the neural folds appearing slowly, then they rear up and the neural tube snaps shut in the middle, and the closing process runs towards both ends, which eventually seal themselves more gently. In [7] the slime mold fruiting body heaves itself slowly off the ground, then accelerates and rises rapidly up its stalk, then slows down, and eventually the knob at the top gently disappears.

We shall now show, by computing the energy released, that this is the normal pattern for a morphogenetical movement arising from a secondary wave.

Note that in the films it can also be observed that the gastrula gives a final heave before the blastopore closes, and the slime mold fruiting body gives a hiccup halfway up; but these are subsidiary frictional effects that we shall explain later.

Assuming that the primary wave begins at time t_1 and ends at time t_2 , let

$V(t)$ = speed of primary wave at time t , $t_1 < t < t_2$;

$A(t)$ = area of wave-front at time t , $t_1 < t < t_2$;

$e(\tau)$ = rate of energy-release by a cell at time-interval τ after the primary wave has passed it. We may suppose that $e(\tau) = 0$ outside an interval $\delta_1 < \tau < \delta_2$, where $0 < \delta_1 < \delta_2$, and where δ_1 is the delay between the primary and secondary waves, and $[\delta_1, \delta_2]$ the period during which energy is released by the secondary effect.

Lemma 1. The total rate of energy release at time t , where

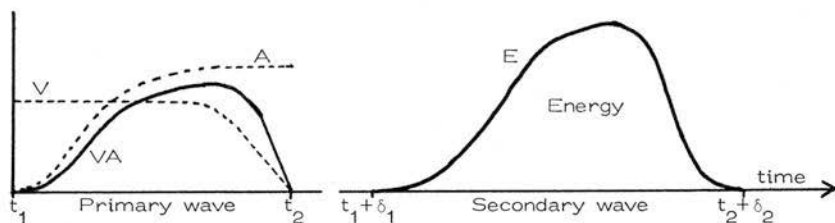
$t_1 + \delta_1 < t < t_2 + \delta_2$, is given by

$$E(t) = \int_{t-\delta_2}^{t-\delta_1} A(\tau)V(\tau)e(t-\tau)d\tau$$

Proof. The number of cells crossed by the primary wave in the interval $[\tau, \tau + d\tau]$ is $A(\tau)V(\tau)d\tau$, and by the time t each of these is still releasing energy at the rate $e(t-\tau)$. Integrating gives the lemma.

In Figure 6 we sketch the qualitative shape of the graph of E . The assumptions on which the sketch is based are as follows: V is initially constant near t_1 , by Corollary 1, and eventually decreases linearly to zero at t_2 , by Corollary 3. If we assume that the primary wave emanates from a point, and that the wave-front initially expands linearly, then we deduce that A starts from zero at t_1 and initially increases parabolically, in other words proportional to a square law. Eventually A becomes constant as the frontier stabilises at t_2 . We assume the secondary effect e starts suddenly at δ_1 , and then decreases linearly to zero by δ_2 . It can be shown, by using the lemma to integrate these assumptions, that E both begins and finishes proportional to cube laws.

Figure 6. Graph of energy released by a secondary wave.



Summarising : this characteristic pattern of energy release during a morphogenetic movement might provide a useful clue that the movement was the secondary effect of an earlier hidden primary wave. George Oster suggests that the energy released by individual cells might be observed by measuring the heat loss microcalorimetrically.

The ripple ahead of a wave. Consider a fixed time t during the primary wave, $t_1 < t < t_2$. The state of the tissue is obtained by lifting the section Sxt of C up onto G in Figure 5. Consider the variation in the state of the cells as the frontier is approached from either side. On the a -side the state is approximately constant, and so the cells are homogeneous, but on the b -side the variation is parabolic as the frontier is approached.

In some cases this phenomenon might be repeated visibly in a secondary wave. For instance, if we had a situation as in §6(ii) above, where the secondary wave was differentiation, then the phenomenon might be visible as a slight ripple ahead of the wave. For example, suppose the expanding frontier of the insect eye were a secondary wave. Then inside the frontier the already recruited eye cells should appear relatively homogeneous, but outside the frontier, the epidermis cells just about to be recruited might show physiological signs of the impending recruitment. In mammalian eyes the effect might appear on the inside of the frontier, because the wave goes the other way.

APPLICATIONS

10.

GASTRULATION OF AMPHIBIA.

For the mathematical reader unfamiliar with gastrulation a recommended introduction is to read Balinsky [1, Chapter 8], and to see the film [11] from the Göttingen film library. The following diagrams are taken primarily from [1,20]. Figure 7 shows photographs of what gastrulation looks like from the outside, while Figure 8 gives diagrams of the main morphogenetical movements going on inside at the same time. Figure 9 shows supporting photographs of sections. Figure 10 gives Vogt's drawings of the 3-dimensional flow of mesoderm cells from the surface into the interior, and Figure 11 the resulting flow of the mesoderm mantle, finishing with its position at the end of gastrulation and beginning of neurulation, with a photograph of the corresponding section. Figures 12 and 13 give Vogt's detailed fate maps. Figures 14 and 15 are diagrams of the subsequent neurulation, leading to the tail bud stage in Figure 16. Figures 17 and 18 show part of the normal tables for newts [9] and frogs [17], giving an idea of the timing involved.

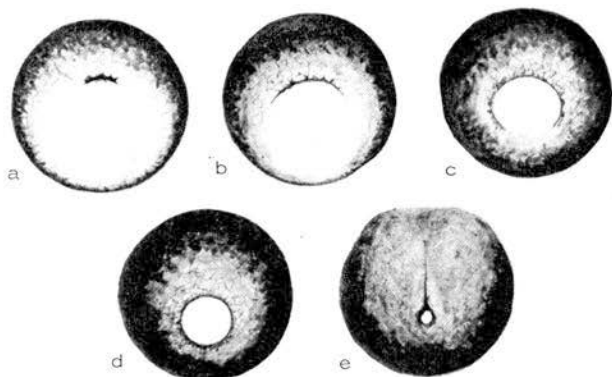


Figure 7 Changes in shape of the blastopore and closure of the blastopore during gastrulation in a frog. From Balinsky [1, p. 187]

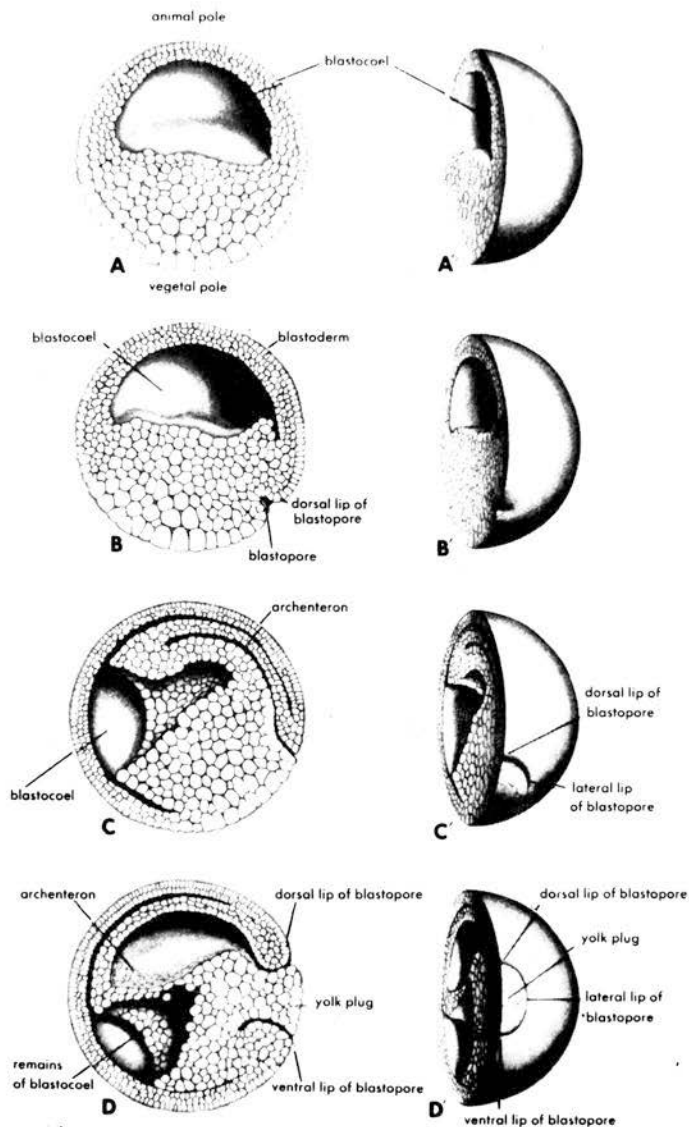


Figure 8 Four stages of development of a frog embryo: A, A', late blastula stage; B, B', beginning of gastrulation; C, C', middle gastrula stage; D, D', late gastrula stage (semidiagrammatic). Drawings on the left represent the embryos cut in the median plane; drawings on the right represent the same embryos viewed at an angle from the dorsal side (A, B, C) or from posterior end (D). From Balinsky [1, p. 191]

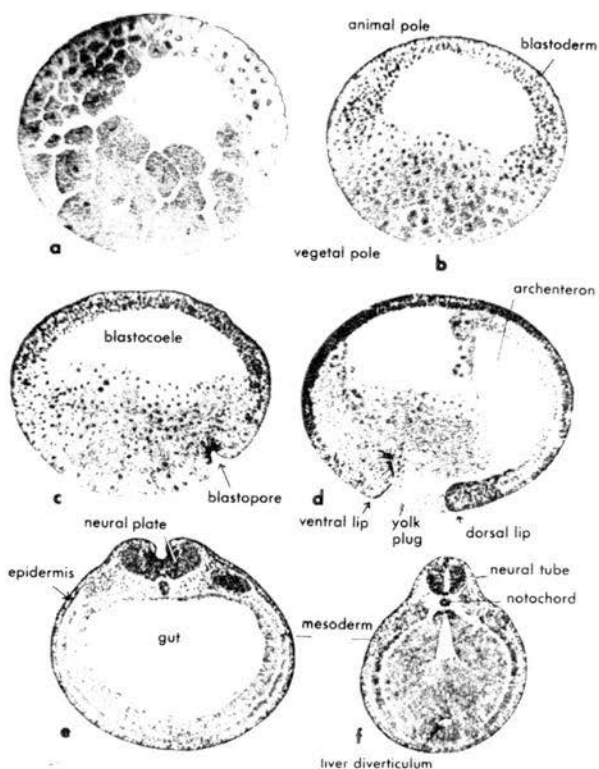


Figure 9 Blastula, gastrulation and formation of primary organ rudiments in the frog. *a*, Late cleavage stage showing difference in the size of the blastomeres; *b*, late blastula; *c*, early gastrula; *d*, middle gastrula; *e*, neurula, transverse section; *f*, transverse section after completion of neurulation. From Balinsky [1, p. 192]

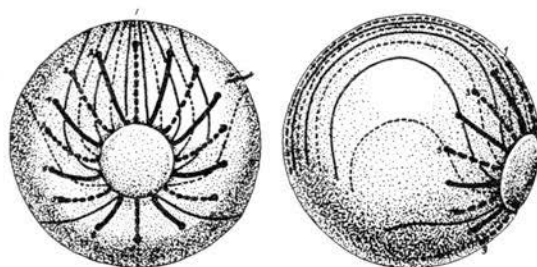
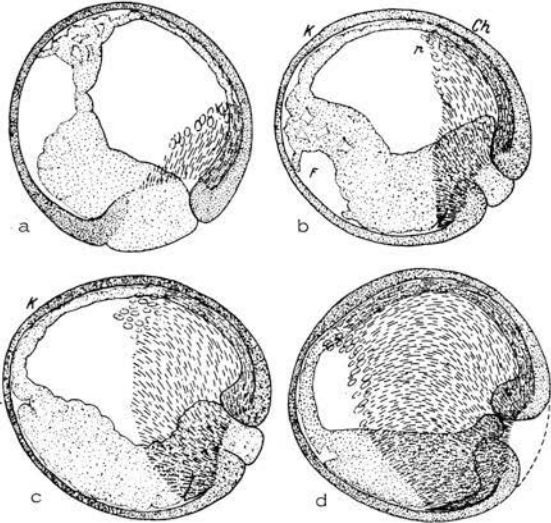
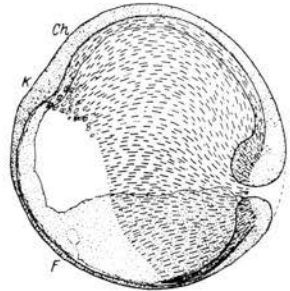


Figure 10 Trajectories of the movements of parts of the marginal zone during gastrulation in amphibians. Thick lines show movements of cells on the surface of the embryo; thin lines show movement of invaginated cells. (From Vogt, 1929.) [20, p. 431] and [1, p. 194]

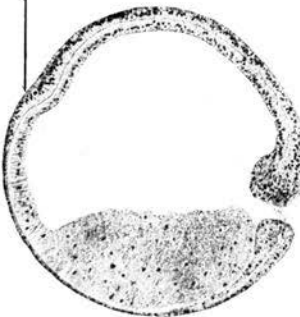
a-d . Pleurodeles. Medianschnitte von Gastrulationsstadien mit rekonstruiertem Mesodermantel. Die Strichelung deutet die Invaginationsrichtung an. Der zellig bezeichnete Mesodermbezirk (p in $\text{\textcircled{D}}$) hat engeren Zusammenhang mit der Urdarmwand und ist stellenweise gescheckt = prächordales Mesoderm. K Vorderende des Medullarbereichs. Ch präsumptives Vorderende der Chorda, zugleich vorderer Abschnitt des kleinstelligen Dachbereichs. F Rest der Furchungshöhle. Der Mesodermantel geht vom Umschlagrand in den Urmundlippen aus, welcher durch den Beginn der Schraffurierung angedeutet ist. — Die Orientierung nach der jeweiligen Schwerkraft zeigt die „Rückdrehung“ des Keimes während des Urmundschlusses.



e Pleurodeles, beginnende Neurula. Medullarplatte deutlich angelegt. Medianschnitt mit rekonstruiertem Mesodermantel. Bezeichnungen wie vorher. Unter dem Kopftomedullarwulst (K) liegt die prächordale Platte und neben ihr das zellig gezeichnete prächordale Mesoderm.



Beginning of neural fold



f Der Medianschnitt, welcher der vorigen Rekonstruktion zugrunde liegt. Ventrale Lippe scharf gesondert vom Entoderm. Urdarmloch in Länge der Medullaranlage entodermfrei, beginnende Ordnung der Zellen im Oberdachbereich. Verdickung der entodermalen Vorderwand des Urdarmes unter entsprechendem Vortücken der prächordalen Platte (vgl. Abb. 36 und 40. — Vogt 40.)

Figure 11 a-e: Flow of the mesoderm mantle over the archenteron roof during gastrulation. f: Photograph of a corresponding median section at the end of gastrulation and beginning of neurulation. From Vogt [20, p. 520].

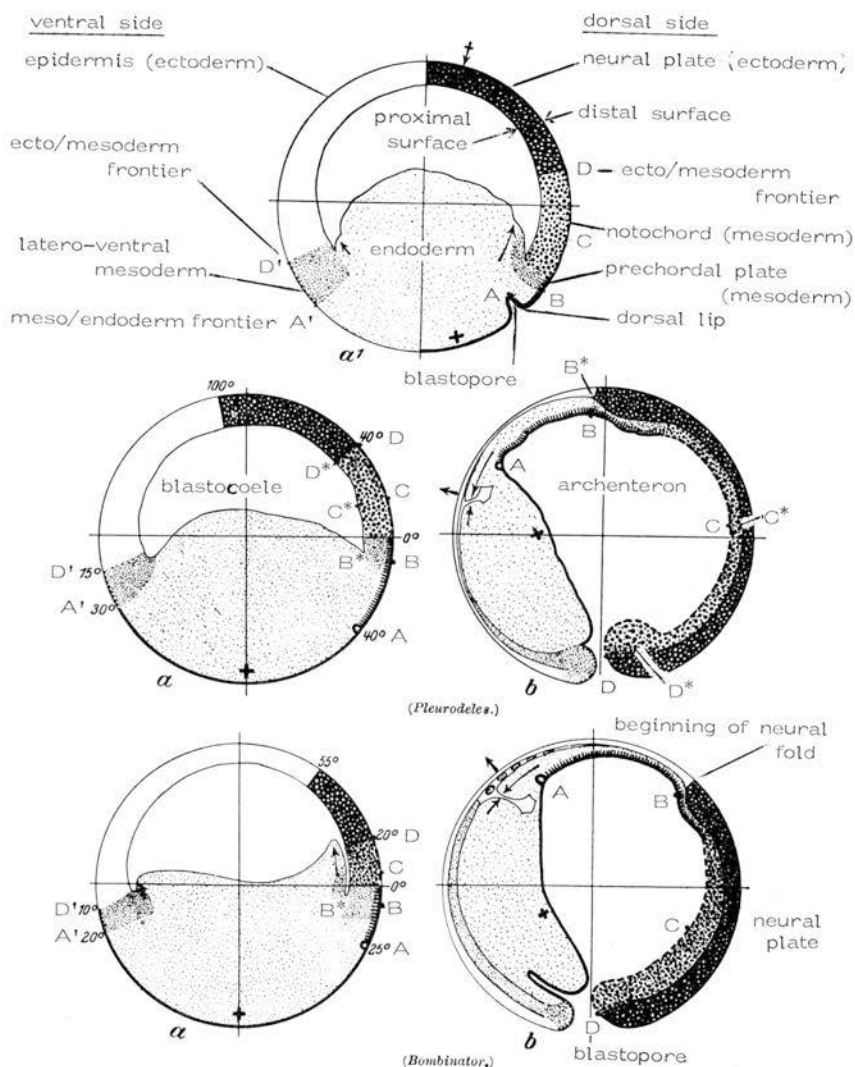


Figure 12. Diagrammatic median sections and fate maps of amphibian embryos (a) before, (a') during, and (b) after, gastrulation. The top three diagrams are for the newt, and the bottom two for the fire-bellied toad. The plus sign marks the track of the vegetal pole, A the track of the initial invagination point, B the track of the point opposite the boundary point B* of the proximal surface, C the midpoint of the notochord, and D the presumptive ecto/mesoderm frontier. Notice that the final position of B in Fig. (b) determines the boundary of the neural plate and the beginning of the neural fold. From Vogt [20, p. 657].

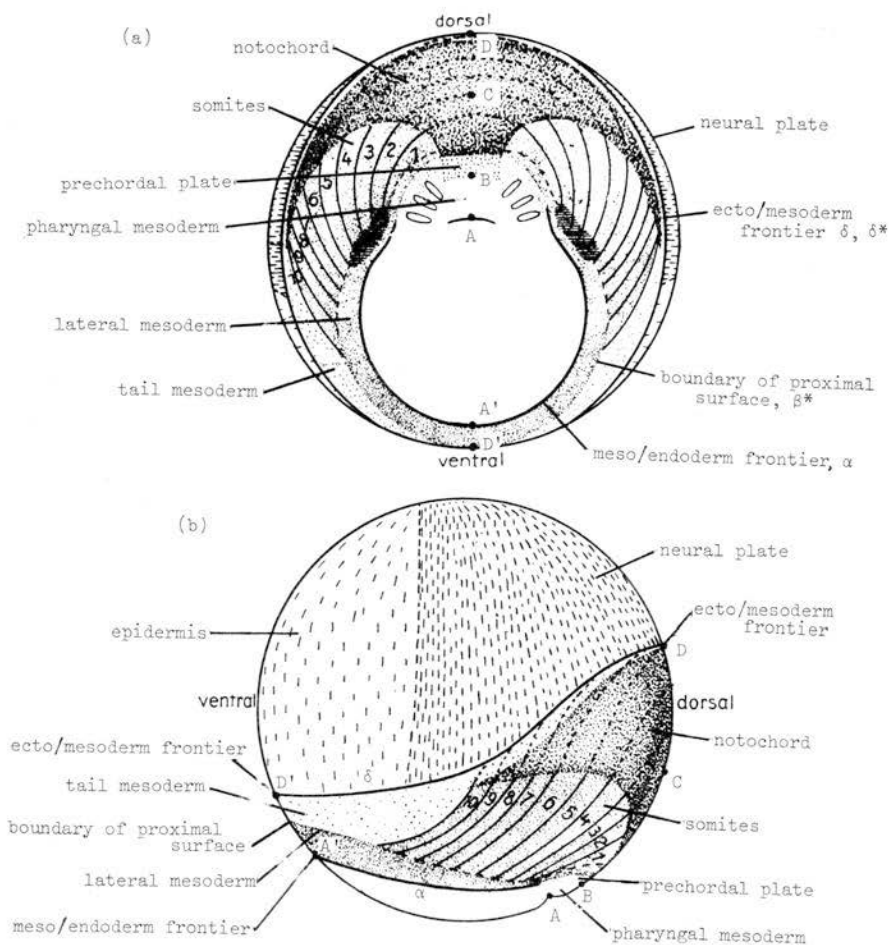


Figure 13. Diagrammatic fate map for urodeles (newts and salamanders) at the beginning of gastrulation, as seen from the outside (a) from below and (b) from the side. The symbols A, B, ... indicate the same as in Figure 12. Invagination is beginning at A, and extends slit-shaped along the short arc through A shown in (a). Notice that the boundaries between presumptive somites are approximately circular arcs centred at A. We have joined these by dotted arcs running through presumptive notochord to indicate the primary wave-fronts. From Vogt [20, p. 392].

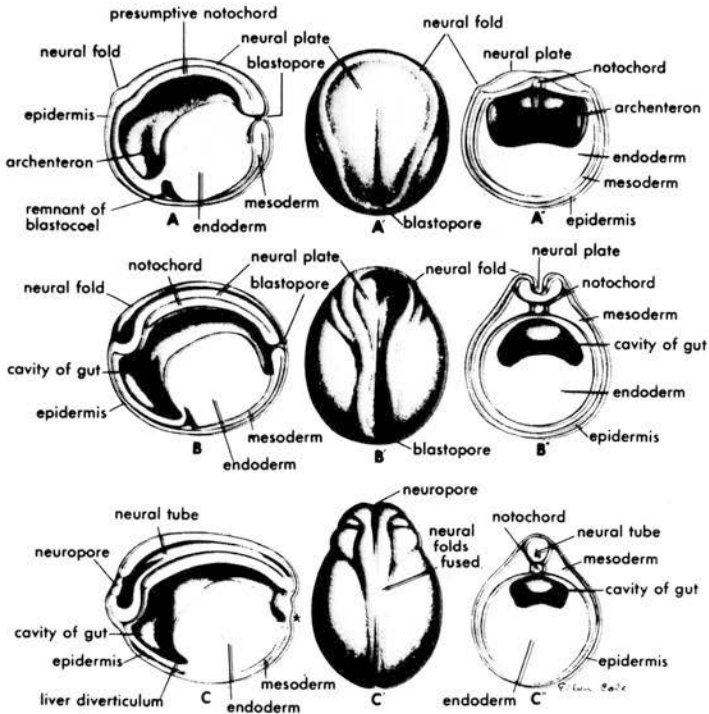


Figure 14 Three stages of neurulation in a frog embryo. The drawings in the middle show whole embryos in dorsal view. The drawings on the left show the right halves of embryos cut in the median plane. The drawings on the right show the anterior halves of embryos cut transversely. A, A', A', Very early neurula; B, B', B', middle neurula; C, C', C', late neurula with neural tube almost completely closed. C shows the blastopore closed; the asterisk indicates the point at which the anal opening will break through. From Balinsky [1, p. 196]

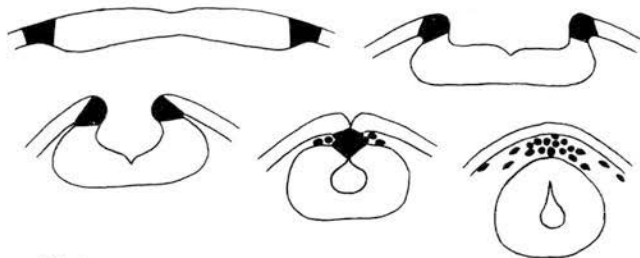


Figure 15 Stages in the formation of the neural plate and neural tube in amphibians. Transverse sections (diagrammatic). Neural crest cells are shown in black. From Balinsky

[1, p. 198]

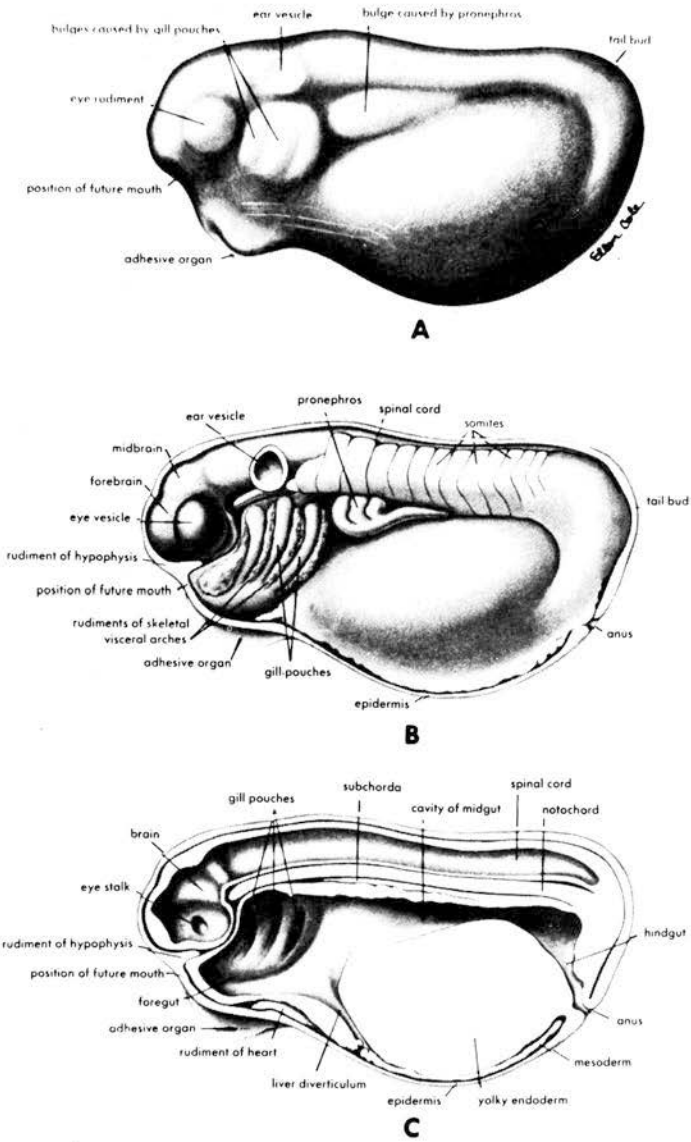


Figure 16. A frog embryo in an early tail bud stage. A, External view. B, same embryo with the skin of the left side removed. C, same embryo cut in the median plane.

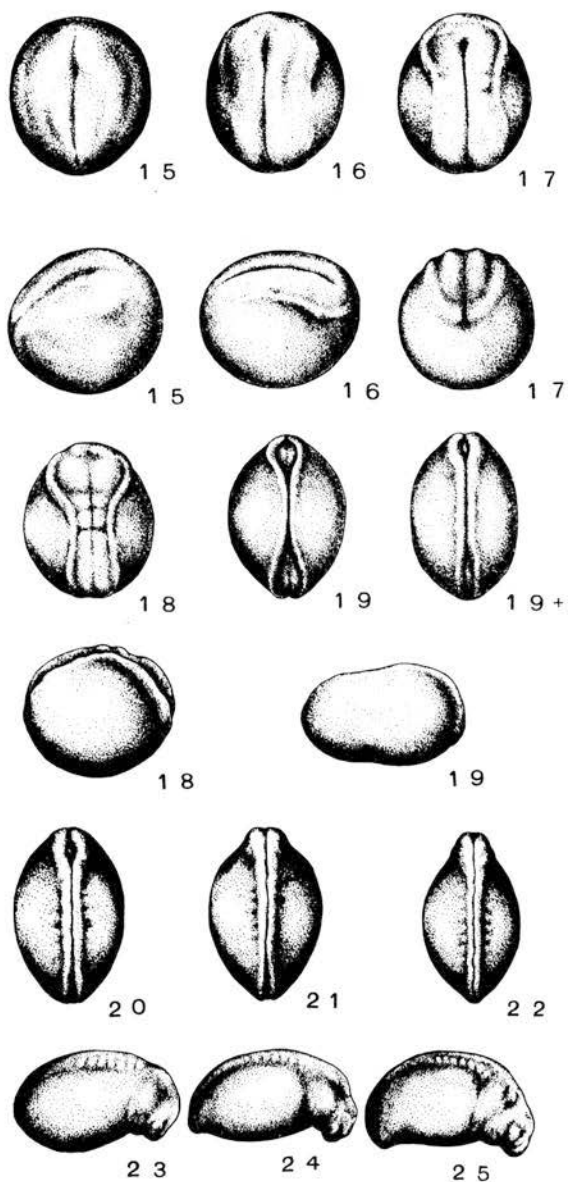


Figure 17. Normal development of the Japanese newt, showing stages of neuralation and formation of somites. From Koyama [9, pp 468-9].

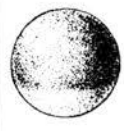
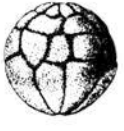
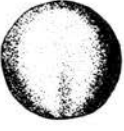
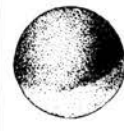


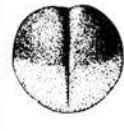





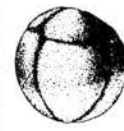

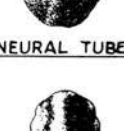


STAGE NUMBER		STAGE NUMBER		STAGE NUMBER				
AGE-HOURS AT 18°C		AGE-HOURS AT 18°C		AGE-HOURS AT 18°C				
1	0		7	7.5		13	50	
		UNFERTILIZED		32-CELL				NEURAL PLATE
2	1		8	16		14	62	
		GRAY CRESCENT		MID-CLEAVAGE				NEURAL FOLDS
3	3.5		9	21		15	67	
		TWO-CELL		LATE CLEAVAGE				ROTATION
4	4.5		10	26		16	72	
		FOUR-CELL		DORSAL LIP				NEURAL TUBE
5	5.7		11	34		17	84	
		EIGHT-CELL		MID-GASTRULA				
6	6.5		12	42				
		SIXTEEN-CELL		LATE GASTRULA				TAIL BUD

Figure 18. Normal development of *Rana Pipiens*.

From Shumway [17, p 143].

Having described the morphogenesis pictorially, we now get down to the main business of discussing the causes. Gastrulation can last from a few hours to a few days, depending upon the species and the temperature (see Figure 8 and Figure 19 (Stages 10-13), and §14 below). During this period there appear to be two local physiological processes going on at the same time, which at first sight seem to have almost the opposite effect at cell level :

- (1) cells flattening,
- (2) cells submerging.

However, whereas process (1) goes on uniformly in all cells of the northern hemispherical shell throughout gastrulation, process (2) does not happen simultaneously everywhere. Our main hypothesis will be that process (2) is secondary wave. Therefore the two processes have radically different global effects upon the shell, as follows :

- (1) Uniform expansion.
- (2) Rolling changes of curvature.

(1) The expansion process. During gastrulation the northern hemispherical shell expands to about twice its area and half its thickness, due to the outer surface cells flattening themselves, and to intercalation amongst the cells of the lower layers, until the shell is only two cells thick [1, Chapter 9]. Moreover this is an active process, because Holtfreter has shown that if the gastrula is prevented from folding inwards then the expansion will force it to fold outwards [21, p.442].

(2) The wave process. We begin with the fact that the northern hemispherical shell differentiates into ectoderm and mesoderm (we are not concerned with the southern hemisphere because it is already different, being yolky, and destined to become endoderm). Therefore, by the theorem, the differentiation will cause an ecto/mesoderm frontier to form, and move as a primary wave. The final position of the wave on the outer surface is a circle δ , whose position can be read from the fate maps (Figures 12 and 13). On the dorsal side δ goes through the point D, and forms the frontier between presumptive neural plate and presumptive notochord, while on the ventral side δ goes through D' and forms the frontier between presumptive epidermis and presumptive tail mesoderm. For the beginning position of the wave we need a hypothesis :

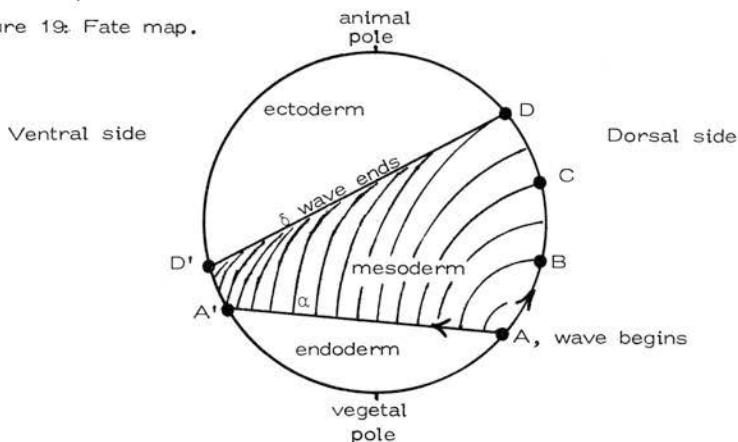
Hypothesis 1. The primary wave begins at the point A on the dorsal side at the bottom of the grey crescent.

It is no accident that A is same point where several hours later the gastrula first begins to invaginate, because the latter is the beginning of the secondary wave (see Figures 7(a), 8(b), 9(c), 12(a₁), 13 and 20).

Let A' denote the point on the ventral side marking the presumptive meso/endoderm frontier, and let α denote the circle on the outer surface through A and A'. From Figures 12 and 13 it can be seen that α is almost a parallel of latitude.

After beginning at A the primary wave then spreads simultaneously in the following directions: firstly east and west round the circle α towards the ventral side, secondly northwards up the dorsal side, and thirdly inwards from the distal (outer) surface to the proximal (inner) surface of the northern hemispherical shell. In Figure 19 we illustrate this on a simplified version of the fate map of Figure 13(b), showing successive positions of the wave on the distal surface.

Figure 19: Fate map.



The extent of travel up the dorsal side, AD, is greater than that on the ventral side, A'D', and it differs in different species. Examples of the latitudes of A and D are as follows:

	Newt	Fire-bellied toad	Axolotl
Wave begins at A :	40°S	25°S	30°S
Wave ends at D :	40°N	20°N	20°N.

The figures for newt and toad are from Vogt (Figure 12), and those for

axolotl are due to Nieuwkoop [12]. Nieuwkoop confirms that a hidden wave of mesoderm determination can be detected in axolotl by the grafting experiment described in §7 above. We return to the timing of this wave in §14 below.

Before we leave Figure 19 there is a small point to be mentioned. We have labelled the entire region between α and δ as mesoderm, whereas in the literature there is some ambiguity as to whether the immediate neighbourhood of the point A should strictly be called mesoderm. More detailed fate maps tend to leave the meso/endoderm frontier undefined near A, as for example in Vogt's fate maps in Figures 12 and 13. However we suggest that the ambiguity may be resolved by considering the 3-dimensional picture rather than merely the 2-dimensional distal surface, and shall treat this in detail when we come to analyse neurulation in §12 below, and in particular in Figure 23.

We now come to our main hypothesis.

Hypothesis 2. There is a secondary wave of cells submerging.

The time delay between the primary and secondary waves is discussed in §14 below. For example in the case of the frog (Figure 18) we estimate the delay is about 16 hours.

Remarks about submerging. Recall that submerging means that the cell decreases its free surface, and increases the proportion of surface in contact with other cells. From the cellular point of view this is a complicated process, but the advantage of denoting it by a single word is that we can then more easily describe the process proceeding as a wave across the tissue. However, at the same time, we must be cautious about the dangers of over-simplification, because different cells may submerge in different ways. For instance the submerging behaviour may depend not only upon the gene systems that have been switched on by the primary wave, but also upon characteristics of the cell membrane, which may in turn depend upon the cell's position.

For example consider the beginning of the secondary wave : this occurs at the point A because by Hypothesis 1 the primary wave began at A. The submerging cells remain attached to the surface, causing the surface to invaginate, and this is the beginning of gastrulation. The submerging cells are known as flask cells because they become elongated, as can be seen in the diagram in Figure 20(a) and photograph 20(b).

Figure 20(a) Median section through the blastopore region of an early gastrula of a newt, showing the cells at the bottom of the pit streaming into the interior. (After Vogt, 1929.)
From Balinsky [1, p. 187]

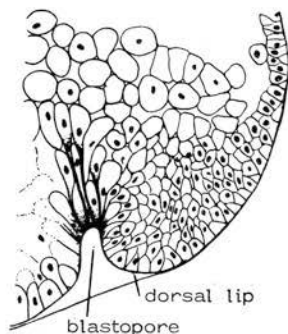


Figure 20(b)

Trit. crist. Gastrulationsbeginn, sagittal, etwas schräg geschnitten. Dorsale Lippe paramedian getroffen. In ihr deutet kleinzelliges Material der inneren Randzone, welches nach innen vorzudringen scheint, die erste Mesodermbildung an. Urmundgrüben ganz von entodermalen z. T. fächerförmigen Zellen gebildet. — Vergr. 50×.

From Vogt [20, p. 510]

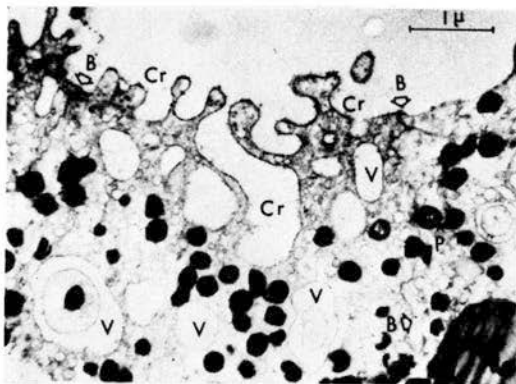


Figure 21 External surface of cells invaginating in the amphibian blastopore (electron-micrograph). Cr, Crypts formed at the surface of the cells; V, vesicles pinched off at the bottom of the crypts; P, pigment granules; B, cell boundaries; Y, yolk platelet.

From Balinsky [1, p. 215]

Possibly the most important part of the underlying biochemical processes in a flask cell may be associated with the contraction near the free surface. By important we mean where the energy is released. The resulting wrinkling of the free surface is shown in the electronmicrograph in Figure 21. This contraction may be responsible for the squeezing of the nucleus and most of the cytoplasm towards the other end of the cell; and this, in turn, may cause the membrane at the other end to bulge out like a balloon, provided it is sufficiently elastic. Hence the characteristic flask-shape. Therefore the amoeboid action of the submerging flask cells streaming into the interior (Figure 20) may not be an active process at all, but merely the passive consequence of the free surface contracting and the fact that the cells have an elastic membrane. All the cells around the circle α (the presumptive meso/endoderm frontier) submerge in the same way.

By contrast, the cells further north on the blastula submerge less dramatically, as can be seen by comparing the development of isolated fractions [1, Chapter 9]. Possibly this is merely because they may have less elastic membranes at the time of submerging. Consequently when the secondary wave hits them they do not balloon out like flask cells, but exhibit more moderate changes of curvature. One might be tempted to say that they submerge more gently, except that the forces involved may be just as powerful.

Again cells on the proximal surface may submerge differently to those on the distal surface, as we shall see when we come to discuss neurulation in §12 below. But the common feature of all submerging cells is that they push and pull on their neighbours. Indeed, as Gustafson and Wolpert [8] have pointed out, the forces that shape the embryo must necessarily originate from cells pushing and pulling on their neighbours, and this is essentially how a secondary wave of submerging cells causes the rolling changes of curvature that produce morphogenesis. Therefore, from the global point of view, the most effective way to describe the results of the secondary wave may be mathematically in terms of changes of curvature. Since curvature will be central to our discussion we digress in the next section to give an elementary mathematical treatment of curvature. But first a remark about the dorsal lip.

The dorsal lip. Mesodermal tissue starts rolling over the dorsal lip about 2 hours after the initial invagination. The main cause appears to be the expansion process (1). Therefore mathematically it should be possible to model the resulting surface flow as the gradient of a potential satisfying Poisson's equation, in other words Laplace's equation with a sink at A and source distributed uniformly over the whole surface north of the circle. This explains why initially there is a flow of tissue from ventral to dorsal side.

The secondary wave of Hypothesis 2 now comes into play, causing the invagination to follow the primary wave both east and west round α , forming the circular lip of the blastopore (shown in Figure 7). The resulting gradient flow on the surface now has a circular sink, as shown in Figure 10. It might be possible to use Poisson's equation to simulate on a computer the time-development of the whole surface flow.

Now consider what happens to the tissue when it rolls over the dorsal lip. First we must explain why the lip itself starts fairly blunt (Figure 20) and then becomes sharper as the flow progresses (Figures 8, 9, 11, 12). We shall show that this is due to the preservation of mean curvature of the shell. We shall then show that the secondary wave, by changing the mean curvature, causes first the concavity of the archenteron roof, and later the formation of notochord and somites during neurulation.

11.

CURVATURE.

Let S be a closed surface, and P a point on S . Let r_1, r_2 be the principal radii of curvature of S at P .

Example 1. If S is a sphere of radius ρ then at each point $r_1 = r_2 = \rho$.

Example 2. Let S be the surface of the mid-gastrula (Figures 7(d), 8(D), 9(d), 10, 11(a), and 18 (Stage 11)), and let P be a point on the dorsal lip. If ρ is the radius of the gastrula, then the average radii of dorsal lip and blastopore are approximately $\frac{\rho}{6}$ and $\frac{\rho}{4}$. Therefore the principal radii of curvature at P are $\frac{\rho}{6}, -\frac{\rho}{4}$ (negative because the

radius of the blastopore points outwards, in other words out of the tissue).

Define the gaussian curvature of S at P to be $\frac{1}{r_1 r_2}$. In example 1 the sphere has positive gaussian curvature, and in example 2 the dorsal lip has negative gaussian curvature.

Define the mean curvature μ of S at P by

$$\mu = \frac{1}{2} \left(\frac{1}{r_1} + \frac{1}{r_2} \right).$$

In example 1 the sphere has mean curvature $\frac{1}{\rho}$ and in example 2 the dorsal lip also has mean curvature

$$\mu = \frac{1}{2} \left(\frac{6}{\rho} - \frac{4}{\rho} \right) = \frac{1}{\rho}.$$

It is no accident that these are the same as we shall prove in Corollary 2.1 below.

Suppose now that S is the distal (outer) surface of a shell of thickness ϵ , with proximal (inner) surface S^* . We have in mind the northern hemispherical shell of the blastula (Figures 8, 9, 12). Define the excess e at a point P of S by

$$e = \frac{\text{incremental area of distal surface}}{\text{incremental area of proximal surface}}$$

Lemma 2. $e = 1 + 2\epsilon\mu$.

Proof. Let O_1, O_2 be the two principal centres of curvature of S at P , corresponding to the principal radii of curvature r_1, r_2 . Let a_1, a_1^* be two small arcs, centred at O_1 and subtending the same angle at O_1 , on the distal and proximal surfaces, respectively, and therefore of radii $r_1, r_1 - \epsilon$. Therefore

$$\frac{a_1}{a_1^*} = \frac{r_1}{r_1 - \epsilon} = \left(1 - \frac{\epsilon}{r_1} \right)^{-1}.$$

Similarly let a_2, a_2^* be two small arcs perpendicular to a_1, a_1^* , centred at O_2 . Then $a_1 a_2, a_1^* a_2^*$ are incremental areas on the distal and proximal surfaces, respectively.

Therefore

$$\begin{aligned} e &= \frac{a_1 a_2}{a_1^* a_2^*} = \left(1 - \frac{\epsilon}{r_1} \right)^{-1} \left(1 - \frac{\epsilon}{r_2} \right)^{-1} \\ &= 1 + \epsilon \left(\frac{1}{r_1} + \frac{1}{r_2} \right), \text{ neglecting } \epsilon^2, \\ &= 1 + 2\epsilon\mu. \end{aligned}$$

This completes the proof of lemma 2.

Corollary 2.1. If ρ , ρ_1 , ρ_2 are the radii of gastrula, dorsal lip, and blastopore at any moment during mid-gastrulation, then

$$\frac{2}{\rho} = \frac{1}{\rho_1} - \frac{1}{\rho_2}.$$

Proof. Consider a piece of mesodermal tissue as it rolls over the dorsal lip. Since the expansion process (1) is uniform and gradual, and since the secondary wave (2) has not yet hit this piece of tissue, we may assume that both the excess, e , and the thickness, ϵ , remain approximately constant* as it rolls over the lip. Therefore by the lemma the mean curvature μ must also be constant. Before the tissue rolls $\mu = \frac{1}{\rho}$, and as it rolls $\mu = \frac{1}{2}(\frac{1}{\rho_1} - \frac{1}{\rho_2})$. Hence the result.

Remark. As the blastopore closes at the end of gastrulation ρ_2 becomes very small, but ρ_1 cannot become too small, because it must at least be greater than the thickness. Therefore μ must become smaller. Therefore ϵ must become greater. This explains the thickening of the lips as the blastopore closes (see Figure 11(f)). Also furrows are often seen on the surface radiating from the closing blastopore, as the distal surface compensates for its increasing negative gaussian curvature. In particular this is probably responsible for the posterior end of the neurulation furrow (see Figure 14 (A') and §12 below).

The archenteron roof. Once the mesoderm has rolled over the lip and got inside it is called the mesoderm mantle. Its mean curvature has changed from being positive to negative, because the distal surface which used to be the convex side of the shell, has now become the concave side, and forms the archenteron roof (see Figures 8,9,10,11,12,14). What causes the change of sign of the mean curvature is the secondary wave of hypothesis 2, progressing along the distal surface shortly after it has rolled over the lip. As the secondary wave hits each cell it will begin to submerge, but it cannot submerge very far because by this time the mesoderm mantle is only two cells thick. Therefore there is no chance of any dramatic behaviour such as displayed by the initial flask cells, which have the luxury of a thick yolky mass of cells in which to submerge. Therefore the main effect of the secondary wave at this stage will be the reduction of free surface membrane (see Figure 21) in other words a reduction in the area of the

* We take account of the gradual decrease in the thickness, ϵ , due to the expansion process in §14 Figure 29.

distal surface, and this effect will persist while the mesoderm mantle climbs up over the archenteron roof (Figures 10, 11, 12). If the distal surface area is reduced by a factor λ , then the excess, e , will also be reduced by λ . We can therefore compute λ by Lemma 2.

Corollary 2.1 If ρ is the radius of the gastrula and ϵ the thickness of the mesoderm mantle, then

$$\lambda = 1 - \frac{4\epsilon}{\rho} .$$

Proof. By the lemma, before rolling over the lip the excess is

$$e = 1 + \frac{2\epsilon}{\rho} .$$

After rolling the excess now becomes

$$\lambda e = 1 - \frac{2\epsilon}{\rho - 2\epsilon}$$

because the radius of the distal surface is now $\rho - 2\epsilon$, and the mean curvature negative. Therefore

$$\begin{aligned} \lambda &= \left(1 - \frac{2\epsilon}{\rho - 2\epsilon}\right) \left(1 + \frac{2\epsilon}{\rho}\right)^{-1} \\ &= 1 - \frac{4\epsilon}{\rho}, \text{ ignoring } \left(\frac{\epsilon}{\rho}\right)^2, \text{ as required.} \end{aligned}$$

Data on shell thickness. From the photographs of section in Figures 9(d) and 11(f) we estimate $\epsilon = \frac{\rho}{12}$. From this single piece of datum we shall deduce several corollaries, including estimates of diameter of notochord and somites in the next section. For the moment we deduce

$$\lambda = \frac{2}{3}, e = \frac{7}{6}, \lambda e = \frac{2}{3} \times \frac{7}{6} = \frac{7}{9} .$$

Therefore:

Corollary 2.3 The effect of the secondary wave is to reduce the surface area by one third.

The secondary wave is the onset of the change in mean curvature, and we have assumed that this effect comes on relatively sharply. The effect must then last for several hours in order to enable the mesoderm mantle to slide round the archenteron roof (see Figures 9(d), 10, 11). In the case of the frog at 20°C we estimate that the effect must last about 20 hours (see Figure 18 and §14 below). As the effect dies away the mesoderm mantle gradually loses the energy to hold itself in negative mean curvature, and will begin to try and reassert its original positive mean curvature. This can be seen most clearly in Vogt's

careful drawing in Figure 11(d). However the mesoderm mantle is prevented from recovering its original curvature by the enclosing sack of ectoderm. Therefore the entrapped mesoderm will begin to exert counterpressures against the enclosing ectoderm, like a mild spring pressing against the underside of the ectoderm. In fact the pressure is visible from the outside, because the margin of the mesoderm mantle pushes up a ripple in the ectoderm, that can be seen running across the outer surface, and is called the gastrulation wave (see the film [11]).

Another effect of the mild springlike action of the mesoderm mantle is to cause it to tear* away laterally from the endoderm, which by contrast moves passively where it is pushed and pulled by the mesoderm.

A third effect can sometimes be seen towards the end of gastrulation as follows. When the mantle reaches the anterior end (Figure 11(d) and (e)) the increased friction between it and the ectoderm may momentarily interrupt the forward sliding of the mantle, even though at the same time fresh mesoderm will continue to be pushed over the blastopore lip by the expansion process; as a result the blastopore will begin to protrude outwards, pear shaped, until the increased forward pressure on the mantle overcomes the friction, and persuades it to continue sliding forward once more. From the outside, this frictional obstruction and release makes it look as though the gastrula is heaving and gasping for breath (see the film [11]).

A fourth, and probably the most important effect of the pressure of the mesoderm against the ectoderm is that it enables the mesoderm to biochemically induce neural plate in the overlying ectoderm, wherever they touch. Had there been no secondary wave of negative curvature, then the mesoderm mantle, once inside, could not have had the same contact with ectoderm, because, in an attempt to retain its positive mean curvature, it would have had to arch itself away from the ectoderm, leaving cavities, and therefore would not have been able to induce neural plate.

* I am indebted to Klaus Jänich for drawing my attention to the topology of this process.

12.

NEURULATION.

Once the neural plate has been induced then neurulation begins. The main feature of neurulation is the neural plate rolling itself up into the neural tube (Figures 14, 15, 17), which eventually becomes the spinal chord and brain. The rolling up is an active process, because isolated pieces of neural tissue tend to curl up by themselves [1, Chapter 10]. It appears that one of the consequences of induction is to make neural cells submerge, thereby causing the mean curvature of the neural plate to change sign and hence causing the rolling up.

However that is not the only event happening during neurulation because at the same time the underlying mesoderm mantle is forming itself into notochord and somites, and this is what we shall be primarily interested in. The notochord is a long thin cylinder running from front to back just beneath the neural tube (see Figures 14, 16, 26, 27). As it forms it elongates itself and goes rigid, stretching the embryo lengthwise. It is only a temporary organ, and later disintegrates and disappears. However while it exists it plays an important role in providing structural support for the formation of spinal chord and skeleton.

The somites are small masses of mesodermal cells, each with a small cavity inside, arranged in two rows on either side of the neural tube and notochord (see Figures 16, 17, 27). They form sequentially from front to back, until there are about 30 or 40 on either side, depending upon the species. Initially they form fairly rapidly and then slow down towards the tail (in frogs initially one pair every 40 minutes, slowing eventually to 2 or 3 a day). Like the notochord, the somites are temporary organs because the cavities soon disappear, and different parts of the somites develop into the vertebrae, muscles and the connective tissue layer of the skin. However, although temporary, the somites do play an important role in laying down the basic pattern, which the subsequent formation of vertebrae, skeleton and spinal chord makes permanent.

Therefore both the notochord and the somites should perhaps be regarded more as templates rather than organs, because as organs they have no function and soon disappear, whereas as templates they create and donate permanent pattern to the animal. The question arises how is this pattern created? We shall show that it is caused by the same secondary wave that caused gastrulation.

At first sight it might appear a little overambitious to attempt a single explanation for such diverse phenomena during both gastrulation and neurulation and so let us put it more tentatively. In attempting to explain gastrulation we were led to postulate Hypothesis 2 in the last section, namely the existence of a secondary wave - and now this same hypothesis leads us to expect further events in the mesoderm mantle that do seem to occur with the correct timing and geometry during neurulation.

For consider what happens when the secondary wave hits the proximal surface of the mesoderm mantle. It may take some time for the secondary wave to cross the mantle, because when the primary wave originally crossed it the shell was considerably thicker (see Figure 9(a)), with thickness comparable to the distance of travel up the dorsal side. Furthermore there is no reason to suppose that the speed of the primary wave through the shell is necessarily exactly the same as its speed across the surface. In the case of the frog we estimate that the time taken to cross the shell is about 16 hours compared with 24 hours to travel up the dorsal side (see §14 below). The secondary wave will follow with the same timing. Therefore by the time the secondary wave hits the proximal surface, its previous effect of reducing the area of the distal surface may have mostly died away. Therefore the mesoderm mantle will suddenly experience a violent reversal of curvature for two reasons: firstly it will be tending to recover its original positive mean curvature as the effect on the distal surface dies away, and secondly this positive curvature will be drastically reinforced as the reduction effect now hits the proximal surface. The mild spring will be suddenly transformed into a violent spring, and the mesoderm will try to coil outwards underneath the neural plate. However we shall show in Corollary 2.5 below that it is impossible for the spring to achieve its desired curvature. As a result the central piece of mesoderm will rip itself off, and roll up separately flexing itself straight and forming the notochord. On either side individual segments will rip off and curl up into somites. Meanwhile the remaining lateral mesoderm remains unaffected by the secondary wave.

To justify these statements we shall first compute the relevant curvature, and then examine the position to which the proximal surface has been carried by gastrulation in order to explain why lateral mesoderm is unaffected. Next we shall analyse the passage of the secondary wave

across the proximal surface, in order to explain why the notochord forms and elongates, and why the somites form sequentially from front to back. Then in §13 we explain why the somites are segmented, how their size regulates to the size of the embryo, and why the later somites are smaller and slower to form. Finally in §14 we discuss details of timing. First the curvature; let ρ denote the radius of the embryo.

Corollary 2.4. When the secondary wave hits the proximal surface the mesoderm mantle will attempt to take up a positive mean curvature equal that of a sphere of radius $2\rho/9$, or a straight cylinder of radius $\rho/9$.

Proof. By the data on shell thickness in §11, the original excess before gastrulation was $e = \frac{7}{6}$. By Corollary 3 the surface reduction factor imposed by the secondary wave is $\lambda = \frac{2}{3}$. Therefore applying the reduction factor to the proximal surface, the new excess is

$$\frac{e}{\lambda} = \frac{3}{2} \times \frac{7}{6} = \frac{7}{4}.$$

Therefore by the lemma

$$2\epsilon\mu = \frac{e}{\lambda} - 1 = \frac{3}{4}.$$

By the data $\epsilon = \frac{\rho}{12}$, and so

$$\mu = \frac{3}{8\epsilon} = \frac{9}{2\rho}.$$

Therefore if the mesoderm were able to assume spherical form, of radius ρ_1 , then

$$\rho_1 = \frac{1}{\mu} = \frac{2\rho}{9}.$$

Alternatively if the mesoderm were able to assume straight cylindrical form, of radius r , then the other principal radius of curvature would be ∞ , and so

$$\mu = \frac{1}{2} \left(\frac{1}{r} + \frac{1}{\infty} \right) = \frac{1}{2r}.$$

Therefore

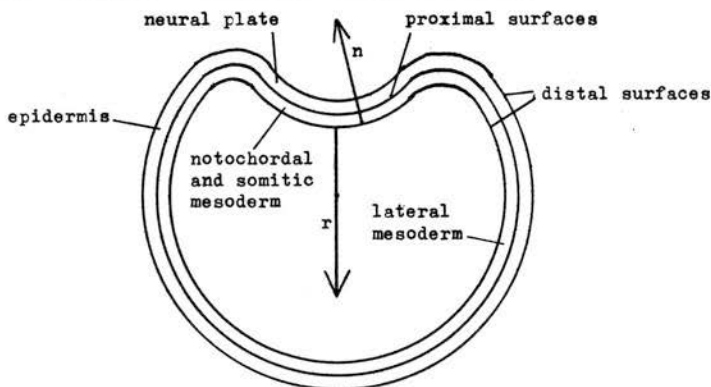
$$r = \frac{1}{2\mu} = \frac{\rho}{9}, \text{ as required.}$$

Corollary 2.5. The mesoderm mantle is unable to achieve the required positive mean curvature without tearing itself.

Proof. Suppose, on the contrary, that it does achieve required mean curvature without tearing itself; then we shall prove a contradiction. When the neural plate begins to roll up, it has negative gaussian curvature (see Figures 13(B') and 17 (Stage 18)), similar to the dorsal lip (§11 Corollary

2.1). However this time the situation in the underlying mesoderm mantle is aggravated on the one hand by the fact that it is trying to achieve an even greater positive mean curvature, and on the other hand by the neural plate getting in the way preventing it from curving so much. More precisely, let n denote the radius of the transverse section, and r the radius of the median section, at a point on the dorsal side of the distal surface of mesoderm, as illustrated in the sketch of the transverse section in Figure 22.

Figure 22. Diagram of transverse section of neurula.



Let ϵ denote the thickness of mesoderm, as before. Initially the overlying neural plate will have approximately the same thickness (see the photograph in Figure 11f), but as it rolls up it becomes thicker. Therefore $n \geq 2\epsilon$. The principal radii of curvature are n , $-r$ (negative because r points out of the mesoderm), and so the mean curvature is

$$\mu = \frac{1}{2} \left(\frac{1}{n} - \frac{1}{r} \right).$$

But the required $\mu = \frac{3}{8\epsilon}$, by the proof of Corollary 4.

Therefore

$$\frac{1}{n} = \frac{3}{4\epsilon} + \frac{1}{r} \geq \frac{3}{4\epsilon};$$

therefore

$$n \leq \frac{4\epsilon}{3}.$$

But this contradicts $n \geq 2\epsilon$, and therefore completes the proof of Corollary 2.5.

A consequence of Corollary 2.5 is that when the curvature or shear forces in the mesoderm become greater than the cohesive forces between neighbouring cells, then the mesoderm mantle will tear itself. In order to explain where the tears will occur, and in what order the torn pieces will curl up, we shall need to examine the way in which the secondary wave travels over the proximal surface. But before we do this, we make a couple of remarks about the validity of Corollary 2.4.

Remark 1. When the notochordal-mesoderm tears free and rolls itself up to form the cylindrical notochord, we might expect its diameter to be $\frac{1}{9}$ that of the embryo, from Corollary 2.4. In fact this is sometimes quite a good estimate (see Figure 26). When the somites form, although they are not exactly spherical in shape, their cross-section is roughly twice that of the notochord, which again is in agreement with Corollary 2.4.

Remark 2. In using the value $\lambda = \frac{2}{3}$ in the proof of Corollaries 2.4 and 2.5 we have implicitly made the rather drastic assumption that cells on the proximal surface submerge in the same way as those on the distal surface, or, more precisely, we have assumed that the effect on the curvature is the same. In fact they may submerge quite differently. However we are justified in using the term submerging, because Mookerjje, Deuchar and Waddington [21, p.451] have shown that the area of contact between cells does increase considerably when the notochord forms. Our implicit assumption is that this local effect imposes the positive mean curvature, which not only forces the cylinder to have small radius, but also to be straight (because if $r < \infty$ in Corollary 2.5, this reduces the mean curvature). Therefore it is the secondary wave of submerging cells that initially forces the notochord to be stiff. Later when the secondary effect wears off the notochord retains its stiffness by the cells vacuolating (that is swelling with fluid) making the whole cylinder turgid, and by the secretion of a thin supporting sheath [21, pp. 262, 450].

The proximal surface, P. Recall that in §10 above we defined the two circles α and δ on the distal surface of the blastula to be the presumptive meso/endoderm and ecto/mesoderm frontiers, respectively. Let X denote the solid piece of blastula bounded by the planes through α and δ . Then X is shaped like a lop-sided barrel, as shown in

Figure 23(X). The surface of X comprises 5 regions as follows :

- (i) The annulus at the top bounded by circles δ , δ^* , to which presumptive ectoderm is attached (shown cross-shaded).
- (ii) The disk at the bottom bounded by circle α , to which the rest of the presumptive endoderm is attached (shown shaded).
- (iii) The inside yolky surface is a disk bounded by the circle β^* , which we have drawn flat for convenience, but which in fact is usually curved (see Figures 9, 12).
- (iv) The distal surface is the outer curved annulus bounded by α and δ , and containing the points ABCD on the dorsal side.
- (v) The proximal surface P, which we define to be the inner curved annulus bounded by β^* and δ^* , containing the points $B^*C^*D^*$ on the dorsal side. Notice that we do not include the yolky surface (iii) in the definition of proximal surface.

The presumptive meso/endoderm frontier lies in the interior of X, and is approximately the conical annulus bounded by α and β^* (shown dotted); this frontier is already predetermined by the boundary of the yolky region of the egg before cleavage. Therefore the sides of the barrel form the presumptive mesoderm, and the floor of the barrel, which is yolky, forms part of the presumptive endoderm, the rest of which is attached to the bottom disk.

Gastrulation turns the barrel inside out, into the bottle-shape Y shown in Figure 23(Y). For diagrammatic simplicity we have not shown all the details of the endoderm part of Y, for instance where it has been torn away from the mesoderm (along the lateral parts of the dotted annulus), but the shape of the mesoderm in Y is sufficiently accurate for our purpose. Let Q denote the new position of the proximal surface after gastrulation, now lying on the outer surface of Y between the circles β^* and δ^* , and containing the points $B^*C^*D^*$. The main point we wish to emphasise is that whereas the distal surface ABCD, between α and δ , now stretches round nearly the whole embryo, the proximal surface Q only stretches over less than half as much. It is the proximal surface that induces neural plate (and not the yolky surface because that is endoderm), and therefore the position of the former determines the latter (see Figure 12). This explains the pear-like shape of the neural plate (see Figures 14, 17, 18, 23). This also answers a question of Waddington

Figure 23. Mesoderm mantle before (X) and after (Y) gastrulation.

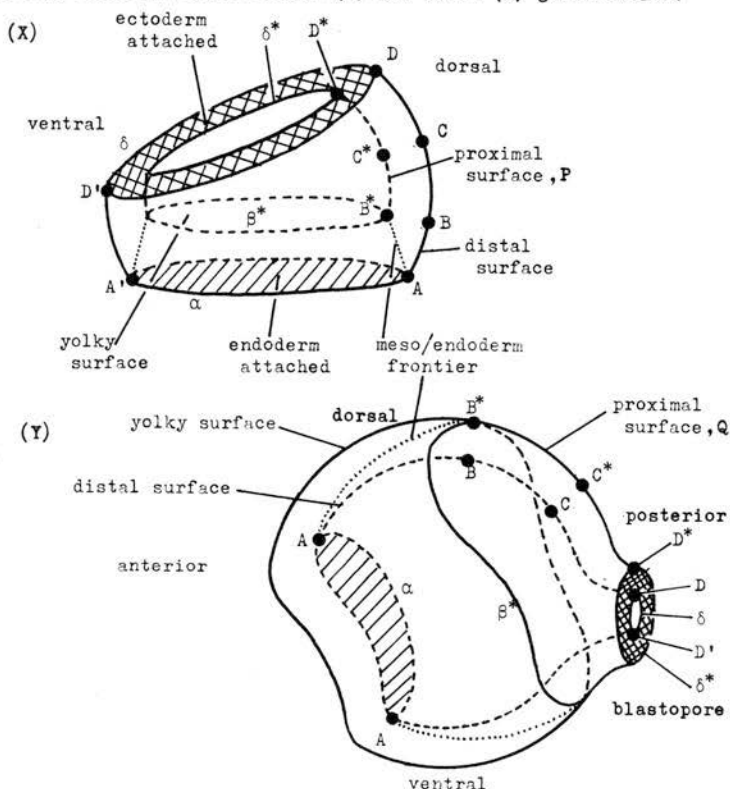
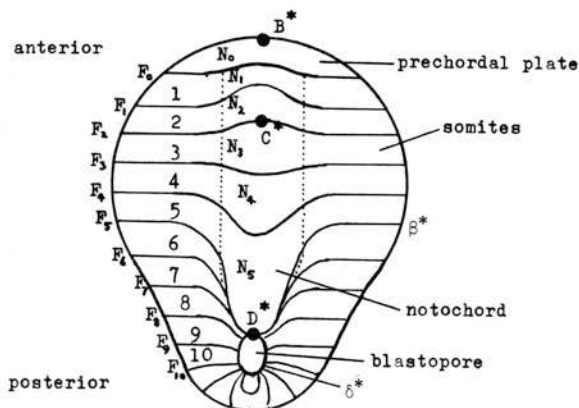


Figure 24. Secondary wave fronts on the proximal surface Q.



[21, p. 467] as to why the neural plate is broad at the front and narrow at the back (which previously seemed to be a paradox on the assumption of an evocator diffusing only from the dorsal mid-line).

The secondary wave-front on Q. Figure 24 sketches the successive wave-fronts as the secondary wave travels across the proximal surface Q, looking down on Q from the dorsal side. Quantitatively we should expect this picture to differ slightly for different species, but to justify the picture qualitatively let us work from Vogt's detailed fate maps for newts in Figure 13, as follows.

Consider first the passage of the primary wave, which travels through only the presumptive mesoderm, that is the sides of the barrel X. The primary wave starts at A, by Hypothesis 1, and travels up the distal surface to the circle δ ; at the same time it penetrates through the shell, first hitting the proximal surface P at B^* , and then travelling up P to δ^* . Let F be the family of curves on P representing the primary wave-fronts, in other words the successive positions of the primary wave.

Now turn to the fate maps of Figure 13. Admittedly these are drawn on the distal surface showing mesoderm as the annulus between α and δ , but by projecting radially from the centre of the embryo, we can identify P with the subannulus consisting of presumptive notochord, prechordal plate, somites and tail mesoderm. P does not contain the complementary subannulus consisting of lateral and pharyngeal mesoderm.

Since the primary wave starts at A, it is reasonable to assume that, near A at any rate, F is approximately the family of semicircles concentric with A. Now the boundaries between presumptive somites are also circular arcs concentric with A, and as we shall explain in the next section this is no accident. But for the moment let us postpone the question of why, when the somites form, they segment themselves along secondary wave-fronts, and merely assume at this stage that the presumptive somite boundaries lie along primary wave-fronts.

Let F_n denote the wave-front separating somites n and $n + 1$. The complete wave-front F_n is a semi-circle, whereas the somite boundaries consist of two arcs, one on either side, and so in order to draw F_n we must join the latter by an arc running through the presumptive notochord, which we have indicated by a dotted line in Figure 13. These dotted lines do not appear in Vogt's original fate map because as opposed to what

happens in the somites, the secondary wave-fronts do not manifest themselves physically in the formation of notochord, for reasons which we explain in the next section. Therefore the notochord when it forms appears as homogeneous and unsegmented.

Now consider the effect of gastrulation upon the wave-fronts. Gastrulation turns the barrel X inside out into Y , and induces a diffeomorphism

$$h: P \rightarrow Q.$$

The family $\{F_n\}$ of primary wave-fronts in P is mapped by h onto the family of secondary wave-fronts in Q .

Lemma 3. The secondary wave-fronts are as shown in Figure 24.

Proof. We examine the diffeomorphism h . Metrically the most striking feature of h is that it shrinks the large circle δ^* into the small blastopore. At the same time h expands areas uniformly by the expansion process (1) of §10 above (mathematically h has constant Jacobean). Therefore since h shrinks the neighbourhood of δ^* longitudinally it must compensate by expanding latitudinally. Therefore the presumptive notochord in P is mapped by h into an elongated roughly rectangular strip on the dorsal side of Q , shown by dotted lines in Figure 24. Meanwhile we know from experimental observation (see Figures 16, 27) that the boundaries between somites are mapped onto lateral lines in Q , which are shown in Figure 24 with the same numbering as in Figure 13. There remains to justify the extension of the wave-fronts into the notochord region in Q .

Since F_0 is near to β^* in P , the same is true for F_0 in Q . Since F_5 touches δ^* at D^* in P , the same is true for F_5 in Q . The region between F_{n-1} and F_n in P consists of the n^{th} somite and a region of notochord, which let us call N_n . An examination of Figure 13 shows that the areas of N_1, \dots, N_5 are approximately proportional to the numbers $1, \dots, 5$. Therefore the same is true for their images in Q because h expands uniformly. The only way to extend the curves F_1, \dots, F_4 so as to divide the rectangular strip of notochord in Q in these proportions is as shown in Figure 24. Finally F_6, \dots, F_{10} all touch δ^* at D , and so the same is true of their images. This completes the proof of Lemma 3.

We can now follow the progress of the secondary wave across the proximal surface and describe its effects.

Lateral mesoderm. First notice that the secondary wave is confined to the proximal surface, because primary wave never crossed the yolky surface, since that was part of the endoderm (see Figure 23). This explains why the thin tongues of lateral and pharyngeal mesoderm remain relatively passive during neurulation.

Prechordal plate. The secondary wave first hits the proximal surface at its most anterior point B^* . Therefore the first piece of mesoderm to be effected is the prechordal plate, lying between B^* and F (see Figure 24). This may be the reason why the anterior neural fold is the first appearance of neurulation (see Figures 11f, 12, 17). Admittedly the neural folds seem to be the initial phase of the neural plate rolling itself up, but Waddington [21,p.476] remarks that it is surprising that the most anterior part of the neural plate should be the last to be induced and yet the first to curl up. We suggest that the prechordal plate may be giving a helping hand underneath. For the mesoderm at this stage certainly maintains close contact with the neural plate (see Figure 11f); and active positive curvature by the underlying mesoderm would have the physical effect of reducing horizontal tension between overlying neural plate cells, providing them with the opportunity of gently increasing mutual contact without the danger of being pulled apart again, and thereby facilitating their columnar formation.

Supporting evidence that the prechordal plate is itself actively curling up is that it begins to tear itself away from the more anterior parts, the pharyngeal mesoderm, the tear occurring along β^* . This tear enables it to achieve both positive gaussian and positive mean curvature, and to surround the fore-brain region (see Figures 14, 16). Therefore it is not subjected to the same severe stresses as the more posterior parts of the mantle, which are prevented by negative gaussian curvature from achieving the desired positive mean curvature without further tearing. Consequently the prechordal plate is the only piece of proximal surface able to retain its integrity as a sheath surrounding the neural tube.

Neural folds. We suggest that as the secondary wave proceeds through the successive wave-fronts F_1, F_2, F_3, \dots it causes the neural folds

to spread from the front round to the sides (see Figure 17 Stage 16). It takes much longer for the folds to reach the back; in fact Sedra and Michael [16] remark upon the fading of the neural folds towards the blastopore at late stages of neurulation.

One can ask the question (c.f. [21, p. 417]): if the secondary wave is also spreading over the interior of the proximal surface (Figure 24), why does the curvature only appear at the edges of the neural plate and not in the middle? An answer is suggested by imagining unrolling a roll of stiff paper and trying to hold it flat: the paper refuses to go flat and retains most of its curvature at the ends, where the bending moment is least. So the enclosing sack of ectoderm may be pressing down passively on the mesoderm mantle at this stage, allowing it only to curl up at the edges (see Figures 14(A''), 15, 17 Stage 17). When the neural plate itself becomes active shortly afterwards, its curvature tends to be the same both in the middle and the edges (see Figures 9(e), 14(B''), 15, 17 Stage 18).

As supporting evidence for this point of view notice Shumway's remarks [17, II], that photographs of sections at this stage show the neural folds flatter than they appear in living specimens, and show the neural furrow shallower. This can be explained as follows: the force pushing up the neural folds is due to the energy released by the submerging mesoderm cells, and when these cells are killed during preparation of slides, this force disappears allowing the elastic resistance of the overlying neural plate to push the folds down again,

Neural furrow. As can be seen from Figure 24 the secondary wave travels much faster over the notochord region than along the somite region. By the time the wave-front F_5 is reached nearly the whole of the former has been covered. One of the first results is that the mesoderm tries to form a tight fold along the dorsal mid-line, similar to the dorsal lip (§11 Corollary 2.1), only more so because of the greater positive mean curvature. In so doing it pinches together the overlying neural plate to form the neural furrow (see Figures 14, 15). This is probably the main reason for the neural furrow, because without mesoderm underneath the furrow tends not to appear [23, p.305], and consequently the neural tube tends to develop with a round cross-section rather than elliptical. Supporting evidence that the mesoderm causes the neural

furrow comes from time-lapse films*, in some of which the anterior end of the furrow can be seen to disappear with a flip, just at the moment when mesoderm underneath would be freeing itself from the neural plate in order to roll up into notochord.

Notochord. We suggest that by the time the secondary wave has reached the curve-front F_5 the mesoderm has begun to tear itself along the dotted lines in Figure 24. What causes the tearing we shall discuss in a moment. But once the tears have been made the mesoderm between the dotted lines is free to roll itself up into a cylinder to form the notochord (see Figures 14, 16, 26, 27), or, more precisely, the cells in that region are free to submerge further by increasing mutual contact [21, p.451], and thereby form the cylinder. The same increase of cell contact causes the cylinder to flex itself straight, which in turn causes it to also tear itself away from the prechordal plate along F_0 , leaving the notochord attached to the rest of the mesoderm only at the blastopore end.

But what causes the tear along the dotted lines? One can avoid the question by merely saying that this region of mesoderm "differentiates into notochord" - but having taken the point of view that notochord and somites are primarily templates rather than organs, we are obliged to look for some more mechanical reason for the tearing and the formation of pattern. We tentatively suggest the the dotted lines may be the locus of maximal bending moment, maximal sheering forces, and maximal weakness of the mesoderm mantle, as follows.

Firstly the bending moment. Figure 25 is a photograph of a transverse section at this stage, showing the notochordal cells submerging away from the proximal surface, inducing positive curvature and bending moment stress in the mesoderm mantle. Meanwhile the increased mutual contact between notochordal cells increases their cohesive strength, and makes the wave-front itself the momentary line of maximal weakness. But from Figure 24 it can be seen that the wave-fronts $F_5, F_6 \dots$ coincide along the posterior end of the notochord boundary, stabilising the line of weakness, possibly sufficiently long for tearing to occur.

* I am indebted to Jack Cohen for showing me his films on axolotl neurulation.

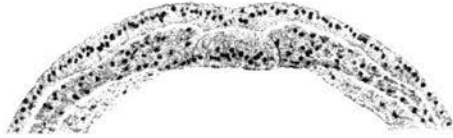


Figure 25. Transverse section of newt neurula showing notochord forming. From Vogt [20, p.544].

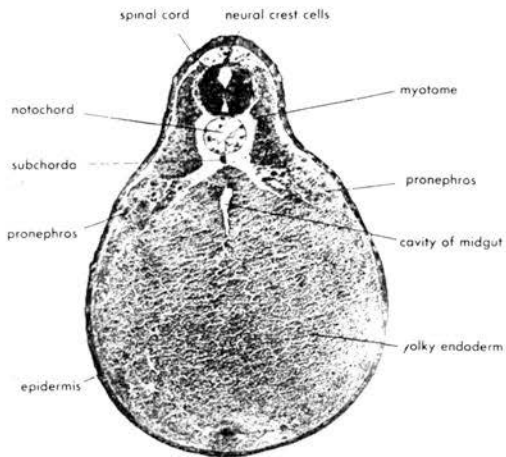


Figure 26. Transverse section of frog embryo showing notochord. By this stage the somites have developed into dermatome, sclerotome and myotome. From Balinsky [1, p.316].

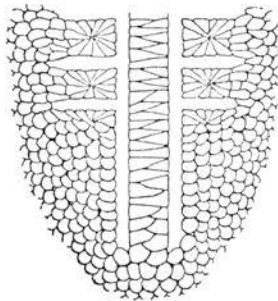


Figure 27. Diagram of cell arrangement in the mesoderm of the newt neurula, showing the developing notochord and somites. From Waddington [21, p.449].

Secondly the sheering forces. The rapid spread of the secondary wave over the posterior end of the notochord region and the consequent submerging will cause the latter to push forward on its anterior part, setting up longitudinal sheering forces. Meanwhile somites 1, ..., 5 will already have segmented, as we explain in the next section, in other words will have torn away from each other laterally along the lines F_0, F_1, \dots, F_5 up to the dotted lines in Figure 24. Therefore each of the first five somites on either side will be attached to the notochord region only along a fragment of the dotted line. This fragment will be subjected to both lateral bending moment and longitudinal sheering force, and, if tearing is to occur, will be the most likely place for the somite to tear away from the notochord.

Admittedly we have suggested rather complicated reasons for tearing, but our description of the forces present appears to coincide with experimental observation. The tearing caused by such forces might be somewhat ragged initially, but then the submerging effect would be an automatic self-correcting device for subsequently rounding up the notochord cylinder and each individual somite (see Figures 16, 27).

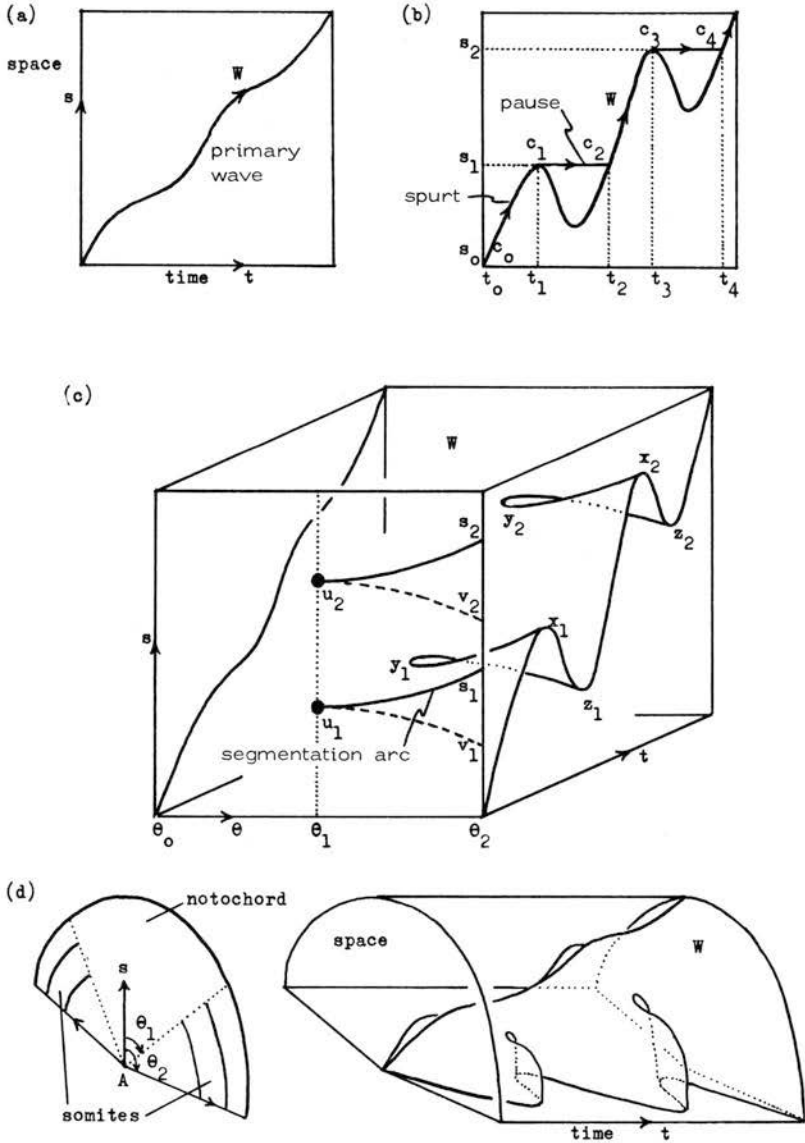
13.

PATTERN FORMATION.

Jonathan Cooke [3,4] has shown experimentally, by surgery, that the somites are determined before gastrulation begins, at just about the same time that our hidden primary wave is beginning to travel up the presumptive mesoderm in the blastula shell. Coupling this with the fact that the presumptive somite boundaries coincide geometrically with the semi-circular primary wave-fronts in Figure 13, it is almost irresistible to conclude that the primary wave is connected with the determination of the somite pattern. Stimulated by Jonathan Cooke, we introduce a new theoretical idea in this section to show how such pattern could be formed. The idea is to combine the primary wave with a clock. There are several ways of combining primary waves with clocks to form regular repeating patterns, and we select one method that seems to give the simplest and most appropriate model for this context.

Suppose that during the passage of the primary wave there is a periodic fluctuation in the levels of concentration of certain chemicals in the cell. Normally the homeostasis and the slow development of

Figure 28. (a) Chemical fluctuations impose a ripple on the primary wave. (b) Fluctuations become dominant. (c) Continuous deformation from (a) to (b) along parameter θ . (d) Segmentation of somites.



a cell might hardly be affected by such fluctuations, but at the moment that homeostatic stability breaks down and the state of the cell suffers a catastrophic jump (as described in the proof of the theorem in §8 and illustrated in Figures 4(b) and 5), then the cell might be exceptionally susceptible to even small fluctuations. For instance fluctuations could affect the threshold, and hence the timing of the jump. More precisely, for any given time t , the state of the fluctuation at that moment will cause the corresponding section of the surface M in Figure 5 to be displaced in the s -direction. As time progresses the section will be displaced to and fro synchronously with the clock. Therefore the clock will impose a ripple on the fold curve of M , and hence upon its projection W in space-time. In other words the clock will impose a ripple upon the track W of the primary wave in space-time, as illustrated in Figure 28.

If the fluctuation becomes dominant, then it will cause the amplitude of the ripple to increase until W is no longer monotonic, as illustrated in Figure 28b. In this case W will no longer represent a continuous wave, because only parts of it will manifest themselves, as is shown by the following lemma.

Lemma 4. If the fluctuation is dominant the the frontier alternately moves in spurts and pauses.

Proof. Using the notation of Figure 28b, during the interval $t_0 < t < t_1$ the arc c_0c_1 represents the frontier moving as a wave from s_0 to s_1 and slowing parabolically to a halt at s_1 . During $t_1 < t < t_2$ the line c_1c_2 represents the frontier pausing at s_1 . During $t_2 < t < t_3$ the arc c_2c_3 represents the frontier moving from s_1 to s_2 , and so on. This follows from tracing the development path of each cell on the surface M lying above space-time, as in Figure 4b.

Remark. Notice that the peaks c_1, c_3 of the ripple manifest themselves, whereas the troughs do not. This is because, near a trough, a cell will already have catastrophically jumped into its new equilibrium state (onto the lower surface of Figures 4a, 5), and so is never able to attain the old stable states represented by the interior of the trough. We therefore say the troughs are silent.

Segmentation of somites. Lemma 4 gives the key to understanding why the somites segment. The primary wave alternately moves in spurts and pauses at the wave-fronts F_0, F_1, F_2, \dots in Figure 13. Therefore

the secondary wave will do the same in Figure 24. Therefore during the spurts from F_{n-1} to F_n all the cells in the n^{th} somite on each side will submerge almost simultaneously, and during the subsequent pause the submerging cells will have time to tear themselves away as a mass from the as yet unaffected region, along F_n . The tear is facilitated by the longitudinal tension set up by the elongating notochord. Once the mass has torn free, the continuing submerging of the cells causes the mass to round up into the somite (see Figures 16, 27).

Having presented the main idea of the model of segmentation we now go on to explain the geometry of the segmentation arcs, and why they end where they do. For as we suggested at the end of the last section, this in turn will determine the line of tear of the somites away from the notochord, and hence determine the boundary of the notochord. Next we shall explain what causes the fluctuations, and the timing of the clock, and finally how the size of the somites is regulated to the size of the embryo. But first we must go back to the theory for a moment.

Segmentation of arcs. Let θ be a parameter representing increasing dominance of the fluctuation over the primary wave, in other words increasing amplitude of the ripple. Figure 28c shows θ drawn as a third dimension, varying smoothly from $\theta = \theta_0$, where the fluctuation is subservient as in Figure 28a, to $\theta = \theta_2$, where the fluctuation is dominant as in Figure 28b. The critical value $\theta = \theta_1$, shown by the dotted line is where dominance first appears, in other words where the graph first becomes non-monotonic.

Now suppose that θ is realised as a second space-dimension perpendicular to s . Then in Figure 28c the (s, θ) -plane nearest the eye represents 2-dimensional space, and the cube represents 3-dimensional space-time. The surface W represents the path of the primary wave in space-time. Generically the projection of W onto the space-plane will have fold curves $x_1 y_1 z_1$, $x_2 y_2 z_2$ which project onto the* cusps $s_1 u_1 v_1$, $s_2 u_2 v_2$. The cusp points u_1, u_2 occur at the critical value $\theta = \theta_1$.

*Mathematically the cusps here are phenomenologically different from the cusp catastrophe shown in Figures 4a,5. The latter arose solely from the dynamic, whereas the cusps here arise from interaction between the time-axis and the dynamic. More precisely they are second order tangencies of time-axis with the fold surface, a 1-higher-dimensional analogue of the point c_2 in Figure 5, which was a first order tangency of the time-axis to the fold curve.

Lemma 5. The primary wave pauses along the arcs s_1u_1, s_2u_2 . Therefore these become the segmentation arcs of the secondary wave.

Proof. From Figure 28b we saw that pauses began at the peaks c_1, c_3 of the ripple. Therefore in Figure 28c pauses begin along the fold curves x_1y_1, x_2y_2 . The projections of these curves onto the space-plane are the arcs s_1u_1, s_2u_2 , which are therefore the segmentation arcs, as required.

Notice that the fold curves y_1z_1, y_2z_2 represent troughs, which are silent, and therefore their projections u_1v_1, u_2v_2 do not manifest themselves in space. That is why we have drawn them as dashed lines in Figure 28c. Note also that each segmentation does not happen simultaneously, as shown by the following corollary.

Corollary 5.1. Segmentation starts at s_1 and proceeds towards u_1 along the segmentation arc s_1u_1 in the direction of θ decreasing. Then along s_2u_2 , and so on.

Proof. The projection of the fold curve x_1y_1 , into the time axis is in the direction of time increasing.

Geometry of the somites. The above arguments work just as well if (s, θ) are taken as polar coordinates rather than cartesian coordinates, and Figure 28d is the diagram for polar coordinates analogous to Figure 28c for cartesian. The critical value $|\theta| = \theta_1$ is shown by the dotted lines, dividing the region $|\theta| < \theta_1$, where the fluctuation is subservient, from the region $\theta_1 < |\theta| < \theta_2$ where it is dominant. The wave fronts are approximately the circles r -constant, and the segmentation arcs are subarcs given by $\theta_1 < |\theta| < \theta_2$. Comparing with Vogt's fate map in Figure 13 we have a reasonably good qualitative explanation for the geometrical shape of the segmentation arcs between successive somites.

Moreover Corollary 5.1 explains why each segmentation begins at the boundary β^* of the proximal surface and proceeds inwards towards the notochord. Figure 16 illustrates clearly the geometry of the somite mesoderm first torn away along β^* curling up round the neural tube, and then the segmentation proceeding successively from front to back, with each segmentation proceeding inwards towards the notochord. In the case of frogs and toads there is an additional detail that each somite

rotates through 90° as it forms [3], and this may be explained by the inward process of Corollary 5.1 coupled with the forces exerted by the submerging cells.

Boundary of the notochord. What determines the limits θ_1, θ_2 of the segmentation arcs in Figures 13 and 28d? The limit θ_2 is easy to understand because this is determined by the boundary β^* of the proximal surface (see §12). But θ_1 is more subtle, and is interesting because in turn it determines the subsequent boundary between somites and notochord. It is well known that before cleavage the gradients in the egg are dominated by the animal-vegetal axis, for instance the pigment is concentrated towards the animal pole and the yolk towards the vegetal pole. Therefore is not unreasonable to assume that when the primary wave travels in a direction having θ sufficiently small, the animal-vegetal gradient will dominate the fluctuation. Conversely if the egg is relatively homogeneous longitudinally it is not unreasonable to suppose that the fluctuation might dominate when θ is sufficiently near $\frac{\pi}{2}$. These two assumptions imply the existence of a critical angle θ_1 between, which in Vogt's fate map in Figure 13 appears to be initially about 30° . As the primary wave proceeds north and begins to slow down before stabilising the effect of the fluctuation may decrease, and so it is not unreasonable to suppose that θ_1 should be an increasing function of s . Therefore we have an explanation of the halberd-shaped region of presumptive notochord in the blastula.

Fluctuations. What is the most likely cause of fluctuations in the levels of chemical concentrations in a cell? One obvious answer is cell-division. Indeed Paul Weiss [23, p.77] points out that mitosis is a kind of earthquake for a cell, which monopolises its resources, and during which development is temporarily suspended. He observes [23, p.85] that cellular differentiation and multiplication are two processes which, if not strictly mutually exclusive, are nevertheless markedly antagonistic in their tendencies. Therefore it is not unreasonable to suppose that it is the regular cleavage in the blastula that produces the fluctuations that cause the ripple during the primary wave, and therefore subsequently the segmentation during the secondary wave.

This hypothesis looks plausible from the point of view of timing, for, from Figure 18, we observe that the cleavage in the blastula occurs at

intervals of slightly less than an hour in frogs, and according to Cooke [3] this slows down to one every 2 or 3 hours shortly after gastrulation has started. If the time delay to the secondary wave is constant, then we should expect a similar timing in the formation of the somites. And sure enough the early somites form one about every 40 minutes [3], slowing to one every 2 or 3 hours in early tail-bud. In the late tail-bud the final somites slow to 2 or 3 per day, but this effect may be enhanced by an increase in the delay between the primary and secondary waves in the final ventral region.

Note that the above figures may be unreliable in that they refer to different experiments, possibly at different temperatures, but are in sufficient agreement to warrant a careful experimental correlation of the timing of cleavage before and during gastrulation, with the formation of somites during and after neurulation.

Regulation. There are nearly always the same number of somites in individuals of the same species. Therefore the size of each somite must regulate to the size of the embryo. But how does an individual cell know whether it belongs to a big embryo or a little embryo, in order to cooperate with the correct number of neighbours to form a big somite or a little somite? To explain this we need to make an additional assumption, that cells at the beginning and end of the wave have predetermined development. Then :

Lemma 6. The length of each somite is proportional to the length of the embryo.

Proof. Let L be the length of the embryo. The length of travel of the primary wave is proportional to L , but the time taken to travel that length is independent of L , since the development of the end points is predetermined. Therefore the average speed of the wave is proportional to L . But the periodicity of mitosis is independent of L . Therefore the distance travelled by the wave between two mitoses is proportional to L . In other words the length of each somite is proportional to L .

Remark. The above proof appears to be similar to the usual argument for regulation in a standard gradient model for pattern formation, which runs as follows : a longer embryo has a shallower gradient and hence longer somites. However Cooke [3,4] points out that the latter argument gives the wrong answer where comparing somites at

the front and back, whereas our model can give the right answer. For near the end of the wave the gradients become shallower, and therefore in the standard gradient model one would expect the somites to become longer, whereas in fact they become shorter. By contrast in our model the shallower gradients are associated with the wave slowing down parabolically (see §9), and so the distance travelled by the primary wave between mitoses decreases causing the somites to become shorter. It would be interesting if it was possible to measure experimentally the parabolic slowing of the primary wave and the decrease in rate of mitosis, for this would then provide a quantitative prediction for the decrease in size of somites. Another way to test the model is to devise a method for altering the speed of either mitosis or the primary wave, without altering the other, because this would then alter the size and number of somites.

14. TIMING OF GASTRULATION AND NEURULATION.

Our whole theory rests upon the existence of a hidden primary wave. As yet there has been little experimental attention directed towards confirming this wave because the very concept of this type of wave first needed the deep mathematics of catastrophe theory. With this mathematics, our theorem predicts the existence of the primary wave, based only upon the fact that mesoderm differentiates from ectoderm.

Our Hypothesis 1 of §10 that the wave begins at A (See Figures 12, 13 and 23) is based on experimental results of Nieuwkoop [12], that in axolotl there is a hidden wave of mesoderm determination travelling up the dorsal side from A to D. This can be detected by the grafting experiment of §7: if the primary wave has already passed the donor point of the graft, then, by transplanting the graft to the ventral side, the graft will induce a subsidiary invagination after the appropriate delay, as in the classical Spemann-Mangolde experiment [1, Chapter 10]. According to Nieuwkoop the primary wave in axolotl begins at A soon after 7th cleavage, and measured in hours after fertilisation has the following timing.

	<u>Hour</u>		<u>Delay</u>
Primary wave begins at A (7th cleavage)	16	}	20
Secondary wave begins at A (dorsal lip)	36		
Primary wave reaches D	36	}	20
Secondary wave reaches D (blastopore closes)	56		

The fact that the delay at A is the same as the delay at D makes Hypothesis 2 plausible, that the secondary wave does indeed follow the primary wave after a 20 hour delay. (The fact that the secondary wave begins at A at the same time as the primary wave reaches D may be accidental because these points are spacially far apart, and may not occur in other species.)

In different species and at different temperature we should expect different timing. For instance from the normal tables for frogs and toads ([13,16,17] and Figure 18) we deduce the following figures. In each case we have assumed that the hidden primary wave begins at A after 7th cleavage (Stage 7-7½) and reaches D by mid-gastrulation (Stage 11), although of course these assumptions needed to be checked experimentally, as in the axolotl case. Meanwhile the visible secondary wave begins with the dorsal lip (Stage 10) and reaches D with the blastopore closing (Stage 13).

Species	[17] <i>Rana pipiens</i>	[13] <i>Rana sylvatica</i>		[16] <i>Bufo regularis</i>	
	18°C	18°C	15.4°C	10.4°C	25°C
Primary wave begins at A	10	6	*4.7	*11	3.5
Secondary wave begins at A	26	19	24	45	7
Primary wave reaches D	34	24	32	60	10
Secondary wave reaches D	50	36	52	96	14.5
Delay	16	12/13	19.3/20	34/36	3.5/4.5

* Note that the figures for the two lower temperatures for *Rana sylvatica* are measured in hours after the first cleavage, as opposed to the others which are measured from fertilisation.

In each case the delay at A equals the delay at D, well within the tolerance of experimental measurement. These figures suggest that Hypothesis 2 remains plausible under wide variations, and would appear to

justify further experimental work to confirm the existence and timing of the primary wave.

We now turn our attention to neurulation and examine in detail the timing for one species namely Shumway's normal table [17] for the frog *Rana pipiens*, shown in Figure 18. Assuming that the primary wave takes $\frac{2}{3}$ as long to cross the blastula shell as to travel up the dorsal side from A to D, we show that Hypothesis 2 continues to be plausible throughout neurulation as well as gastrulation, from the point of view of timing. In order to be precise let us recap, for the frog, all the assumptions that we have made so far about timing. The letters A, B, etc., refer to the points in Figures 12, 13 and 23.

- (i) The primary wave begins at A after 7th cleavage, 10 hours after fertilisation.
- (ii) The primary wave takes 8, 16, 24 hours to reach points B, C, D, respectively, on the distal surface.
- (iii) The primary wave takes 16 hours to cross the shell to the proximal surface.
- (iv) The secondary wave is delayed 16 hours after the primary.

We can now compute the time in hours after fertilisation when the secondary wave hits each of the different points :

Point	A	B	C	D	B*	C*	D*
Hour	26	34	42	50	50	58	66

In order to compute the changes in area of the distal and proximal surfaces we need two more assumptions :

- (v) The effect of the secondary wave is to reduce the surface area by one third (by §11 Corollary 3), and this effect dies away linearly over 20 hours.
- (vi) The expansion process of the blastula shell (see §10 above) starts at hour 20, and lasts for 30 hours, by which time it has doubled the area.

Note that these six assumptions are guesses based on three pieces of data: firstly the timing of the normal table of Figure 18, which Shumway observes can vary up to 10% in individuals; secondly the data on shell thickness at the end of gastrulation, $\epsilon = \rho/12$, from Figures 9(d) and 11(f); and thirdly the estimate $AB = \frac{1}{3}AD$ from Figure 12, which refers to newts

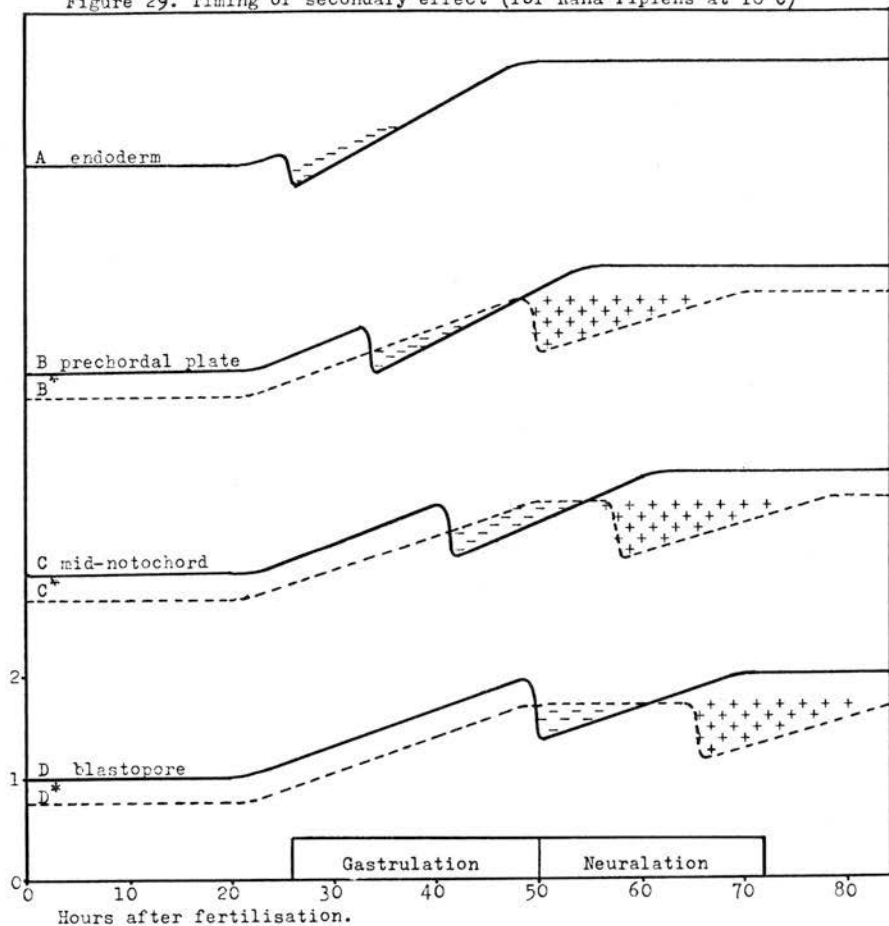
and toads rather than frogs. Therefore our assumptions must be tentative. However we are not claiming quantitative accuracy at this stage, but only verifying quantitative plausibility of the theory.

Using the six assumptions we can now plot against time, in Figure 29, the expansion and contraction at different points of the distal and proximal surfaces of mesoderm mantle, relative to the initial distal area before gastrulation. The curves are drawn for the point A, and the three pairs of points (B,B*), (C,C*) and (D,D*). The continuous curves represent points A,B,C,D on the distal surface, and the dotted curves points B*,C*,D* on the proximal surface. In the case of A there is no corresponding point A* because the proximal surface does not extend that far down (see Figures 12 and 23). In the other three cases we can deduce the excess e and hence the mean curvature μ from Lemma 2.

Since the thickness of the shell after gastrulation is $\epsilon = \frac{0}{12}$, and since it was twice as thick before the expansion process, it must initially have been $\epsilon = \frac{0}{6}$ at the beginning of gastrulation. Therefore by the lemma the excess was initially $e = \frac{4}{3}$. Therefore since we are comparing areas with the initial distal area, all the distal curves must start at 1 while the proximal curves start at $\frac{3}{4}$. (In Figure 29 the scale on the vertical axis indicated refers only to the bottom pair D,D*.)

During the expansion process the distal area doubles, and so the distal curves increase to 2. Meanwhile the thickness halves to $\epsilon = \frac{0}{12}$, reducing the excess to $e = \frac{7}{6}$, and so the proximal curves increase to $2 \times \frac{6}{7} = \frac{12}{7}$. Therefore if there were no secondary wave each pair of curves would be approximately parallel, implying constant positive mean curvature, $\mu = \frac{1}{0}$.

However when the secondary wave hits the distal surface, shown by minus signs in Figure 29, the distal curve suddenly reduces by $\frac{2}{3}$, causing the pair of curves to cross over, and so the mean curvature goes negative, forming (after the initial invagination) the archenteron roof during gastrulation. When the secondary wave hits the proximal surface, shown by plus signs, it is the turn of the proximal curve to suddenly reduce by $\frac{2}{3}$, causing the pair of curves to recross and diverge, and so the mean curvature goes positive, causing the formation of the anterior neural folds, the notochord and early somites. Figure 29 confirms that this takes place mainly during neurulation, between the hours 55 and 70, in agreement with the normal table of Figure 18. Therefore our theory of

Figure 29. Timing of secondary effect (for *Rana Pipiens* at 18°C)

neurulation is plausible from the point of view of timing.

After the primary wave has reached D* it continues to spread round to the ventral side. Therefore the secondary wave, after neurulation, continues to spread ventrally over the proximal surface to form the later somites, eventually extending into the tail bud stages.

15.

EXPERIMENTS TO BE DONE.

Summarising, we have given a unified explanation of both the morphogenesis of gastrulation and the morphogenesis of mesoderm during neurulation, assuming only :

- the shape of the blastula,
- the expansion of the shell,
- the differentiation between ectoderm and mesoderm,
- that differentiation begins at the grey crescent and
- that there is a secondary wave of cells submerging.

The theory admits many testable predictions about timing, and some of the experiments that could be done are as follows :

(i) Verify the passage and timing of the primary wave in different species at different temperatures by graft experiments.

(ii) Find the time that the primary wave takes to cross the blastula shell by splitting the shell and grafting slices of distal and proximal surfaces separately.

(iii) Measure the timing of the secondary waves of curvature change more precisely than the existing normal tables.

(iv) Simulate the morphogenetical movements of gastrulation on a computer, by using Poisson's equation for the expansion process, and the secondary wave of curvature change.

(v) Verify that the archenteron roof, notochord and somites are caused by the same type of cellular phenomenon, by grafting distal and proximal slices onto the surface of endoderm and timing the submerging effect.

(vi) Verify the energy release during the secondary effect by microcalorimetric measurements of heat loss, and by oxygen consumption, in distal cells during gastrulation and proximal cells during neurulation.

(vii) Measure how long the secondary effect lasts, and the speeds of onset and dying away. Compare this with the length of time that the archenteron roof retains its negative mean curvature, the time that a somite preserves its shape, and the time before the notochord cells vacuolate.

(viii) Measure and compare the forces exerted by mesoderm (prechordal plate) and ectoderm (neural plate) in pushing up the anterior neural folds.

(ix) Measure the forces involved in the mesoderm mantle during neurulation, comparing the bending moments and sheering forces with the cohesive forces between cells, in order to test the tearing hypothesis.

(x) Confirm experimentally the pattern of secondary wave fronts on the proximal surface, and verify the timing of the formation of notochord, starting slowly at the front, and accelerating towards the back.

(xi) Compare the timing of mitosis during the primary wave with somite formation during the secondary wave.

(xii) Measure the parabolic slowing of the primary wave and the slowing of mitosis, and predict the rate of decrease of size of somites.

(xiii) By some method of interference, alter the speed of either mitosis or the primary wave, without changing the other, and predict the change of size and spacing of somites.

(xiv) By rotating grafts, and staining, alter the primary wave fronts and relative dominance of the animal-vegetal gradients, and verify the resulting change of presumptive boundaries of notochord and segmentation arcs between somites.

16.

SLIME MOLD CULMINATION.

Slime mold is an interesting species because it normally lives as individual cells, but is also capable of multicellular organisation. When the food runs out the cells stop dividing, and aggregate into a slug-like object called the grex, which first migrates and then culminates into a fruiting body of spores (see Figure 30). When the fruiting body bursts the spores land, and begin life again as individuals.

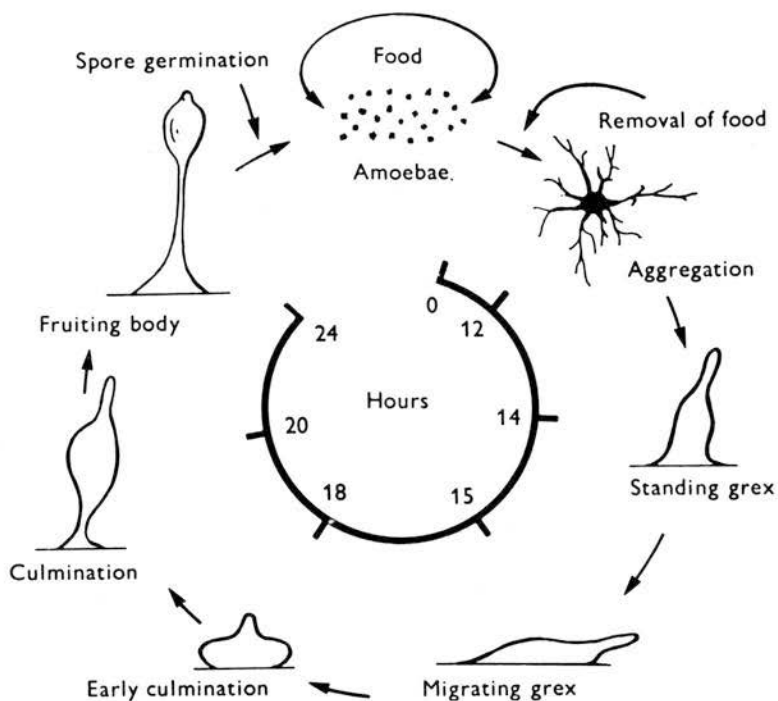


Fig.30 Life cycle of the cellular slime mould *Dictyostelium discoideum*. The times refer to development on Millipore filters (Sussman, 1966). [6, p. 408]

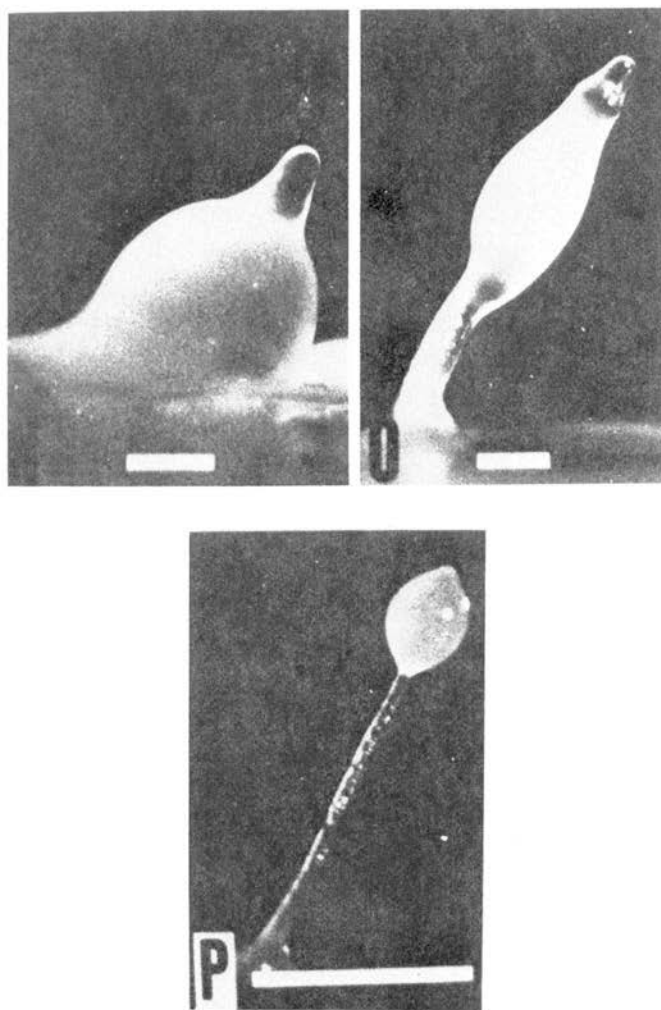


FIGURE 31 N, early culmination about one hour from the end of migration; 0, fruiting body erecting about 3 hours after N; P, eight hours after 0, fruiting body with stalk, spherical spore mass and tip differentiating into spores.

Length marks: N and 0, 1/10mm; P, 1/2mm.

From Robertson [15, p. 58]

For background reading the reader is recommended to read [2,5,6,10, 14,15], and also to see the Göttingen film [7] of G.Gerrisch. Figure 30 is from [6] after M. Sussman, and the photographs in Figure 31 are from [15]. Most mathematical modelling has been concerned with the

aggregation process, but here we shall be concerned with culmination. The morphogenesis of the fruiting body is surprisingly complicated (see [2,5,21]) considering that for most of the time the species lives as single cells. We shall explain the morphogenesis by means of a hidden primary wave, followed several hours later by a secondary wave.

The fruiting body of spores stands on a long stalk made of vacuolated cells, which die after they have performed their function of supporting the fruiting body. Basically there is just one differentiation, which divides the grex into spore and stalk-cells. There may be a slight variation between cells at the foot of the stalk and those higher up the stalk, but since this variation is probably continuous we are justified in saying that all the stalk cells form one type, and all the spore cells another type. Therefore, since there are only two types, this is one of the simplest species to which we can apply the main theorem. The theorem states that the frontier between spore and stalk must move before stabilising, and therefore gives a primary wave. The final position of the wave is the presumptive stalk/spore frontier, which Raper [2,6,14] found experimentally to be about $\frac{1}{3}$ of the way along the grex, the front $\frac{1}{3}$ becoming stalk, and the back $\frac{2}{3}$ becoming spore. As in the case of gastrulation we need two hypotheses :

Hypothesis 1. The hidden primary wave begins at the tip of the grex several hours before culmination.

Hypothesis 2. There is a secondary wave of cells submerging, and then exuding a coating of slime. Note that in this example the submerging cells do not retain any portion of their membrane on the surface of the tissue as in the previous example (Figure 21), but, by amoeboid action towards their neighbours, submerge themselves completely into the interior of the tissue. Therefore the cells do not keep the same topological positions relative to one another.

We shall first discuss the experimental evidence for the primary wave (Hypothesis 1). We shall then explain how the secondary wave (Hypothesis 2) causes culmination and the formation of the fruiting body.

The primary wave. Initially all the cells may have double potentiality, but as the primary wave passes each cell, that cell switches from pre-spore into pre-stalk, and loses spore-potentiality. This is confirmed by the classical experimental results of Raper [14],

and others [2,5,6,10,15] in which pieces were cut off the front and back of the migrating grex, as follows.

Firstly the front pieces culminated into stalks only, with no fruit. We explain this by the loss of spore-potentiality of the entire front third. In the geometrical language of catastrophe theory, the piece is caught on the lower surface of Figure 5, and no part of it can get back up onto the upper surface.

Secondly the back pieces culminated normally, that is to say formed normally proportioned fruiting bodies but took about twice as long to do it. We explain this by assuming all the cells in the back piece to be pre-spore. Consequently the back piece behaves just like the initial grex, only smaller in proportion, so that a new primary wave starts at the front of the piece, and finished $\frac{1}{2}$ the way down the piece. This takes some time, which explains the delay in culmination.

Thirdly, if the front pieces were cut off 24 hours before culmination, then the pieces this time culminated normally. We explain this by assuming that the primary wave had not yet started, and so when it did eventually start it stopped $\frac{1}{2}$ along the piece. This puts an upper bound of 24 hours on the delay between the primary and secondary waves. The experimental data was not oriented towards looking for hidden primary waves, and so is not yet sufficient to determine the delay accurately.

Remark about regulation. These results raise the question of regulation : why does the primary wave always stop $\frac{1}{2}$ along the grex ? We have chosen to take this as an experimental fact but if we wanted to explain this fact, we should have to make an additional hypothesis concerning the nature of the gradient underlying the primary wave. In this example there is strong experimental evidence that this gradient is caused by diffusion, as opposed to the previous example where the gradients in the amphibian blastula were largely inherited from the oöplasmic organisation of the egg before cleavage, rather than caused by subsequent diffusion between the blastomeres. But in the slime mold we start with a homogeneous collection of cells, and the simplest way to create a gradient is by diffusion. And indeed there is evidence of both chemical diffusion, for instance of cyclic-AMP [2,5,6,15], and dynamic diffusion, that is to say periodic activity of cells entrained by signals emanating from a pacemaker [5,15]. In particular the aggregation

process seems to be initiated by chemical signals emanating from one particular cell, towards which the others aggregate, and which then acts as pacemaker. Subsequent periodic surges of movement, entrained by the pacemaker, can be seen during aggregation in the film [7]. Mathematical models of aggregation suggest that the chemical diffusion and the dynamic diffusion are interrelated [15]. When the grex begins to migrate, the pacemaker and source of diffusion is situated at the tip [5], while the sink of diffusion is situated at the tail. Moreover each cell appears to seek its "own" position relative to the diffusion gradient [10], possibly by means of a relative movement induced by comparison between the entrainment clock and some internal clock.

An additional hypothesis sufficient to explain regulation would be that cells at the diffusion source and sink should develop into specific types of stalk and spore cells, independent of the size of the grex. Then diffusion would induce continuity, and impose a different development upon each of the cells in between, according to its relative position. In other words the two ends are regulated by the additional hypothesis, and the final position of the frontier between them is regulated by diffusion.

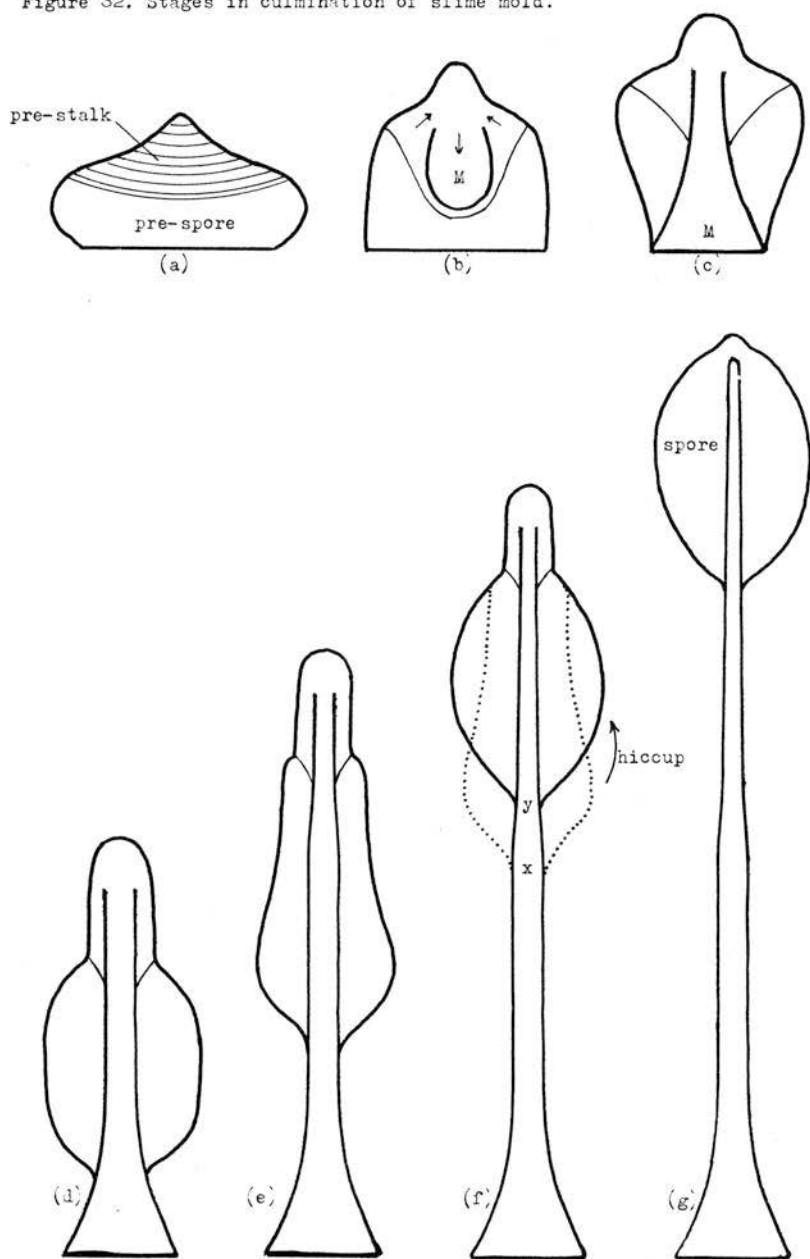
Culmination.

(a) Culmination begins when the grex stops moving and settles down into an onion shape, as shown in Figures 30 and 32(a). The tip of the grex becomes the tip of the onion, and the tail of the grex the base of the onion, and so what were the successive wave-fronts of the primary wave settle into horizontal shallow saucer-shaped layers, roughly concentric with the tip. The wave fronts are sketched (qualitatively rather than quantitatively) in Figure 32(a).

We shall now apply Hypothesis 2. We shall show that when the secondary wave hits this sequence of wave-fronts, from the tip downwards, causing them successively to submerge and exude slime, then this will cause the erection of the fruiting body. Moreover we shall show that this causes several of the various qualitative features displayed in Figure 32. The following sequence of paragraphs (b), (c), ..., (g) refer to the sequence of stages pictured in Figure 32.

(b) The first few wave-fronts have submerged, and form a roughly egg-shaped mass M , that floats below the surface. As each layer submerges it pulls the boundary of the next layer together at the top,

Figure 32. Stages in culmination of slime mold.



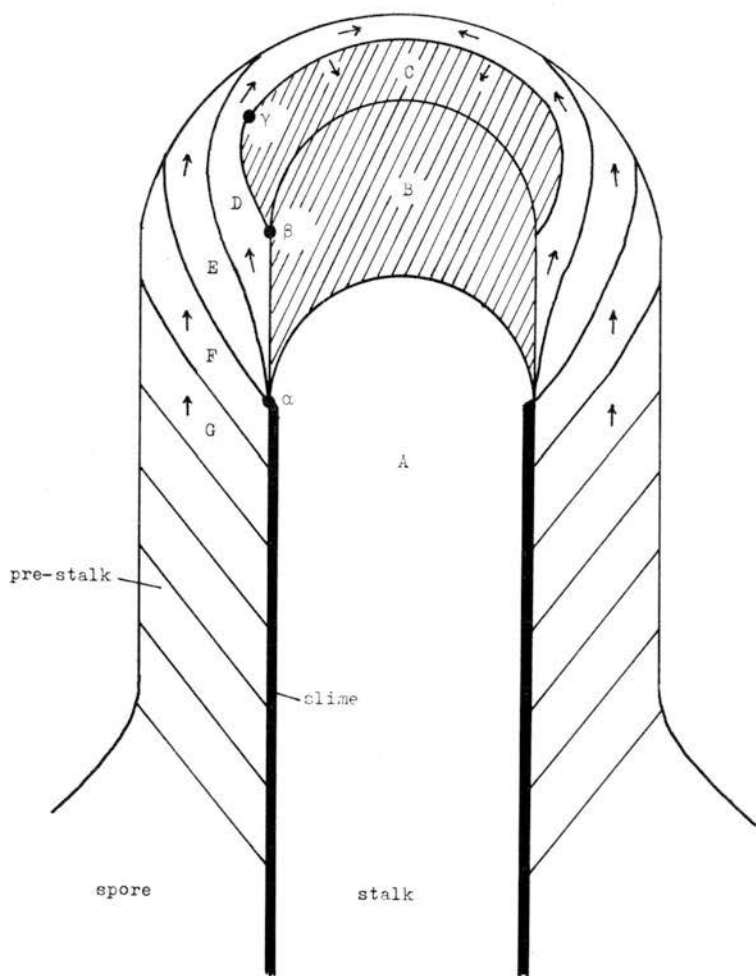
like pulling closed the neck of a sack, enclosing M. The layer that is pulled together at the top, in turn pulls inwards and upwards on its successors, and that causes an equal and opposite downwards push on M. Therefore M begins to move downwards, pushing the centres of the subsequent layers downwards ahead of it, thereby changing them from saucer-shape into cup-shape.

Soon after the cells of M have submerged they begin to exude slime, and so M develops a cup-shaped coating of slime around it, shown by the thick curve in Figure 32(b). This coating of slime facilitates the downwards movement of M, under the downward pressure, and enables M to rupture the subsequent layers, so that they become shaped like the sides of a cup without the bottom. Therefore topologically each layer is first punctured at the bottom, into an annulus, and then pulled together at the top, back into a disk again. This topological puncturing and glueing does not in any way interrupt the passage and timing of the secondary wave, because the latter is merely a set of local events in cells occurring at a fixed time delay after the primary wave.

(c) By this time the downward pressure on M has pushed it clean through the onion to the ground, where it bulges out to become the foot of the stalk. The slime coating eventually dries to form a cellulose tube round the stalk, giving it structural strength. But as yet, especially near the top of the stalk, the slime is still slippery, and so as the secondary wave hits each successive wave-front, causing it to submerge, it cannot adhere to the slimy walls of the stalk, and so is forced to adhere to the top of the stalk. Therefore the stalk starts growing upwards, pulling the rest of the onion with it. Once the latter pulls clear of the ground, it slides easily over the foot because the latter is conical in shape and offers no frictional resistance. Therefore the onion slides into shape (d).

(d) By this stage the characteristic knob has appeared at the top of the growing fruiting body (see the photograph Figure 31), and persists through stages (e) and (f). The structure of the knob is sketched in detail in Figure 33. In order to describe the geometry of what is going on in the knob, we shall assume discrete intervals of time, and have marked in successive wave-fronts accordingly. Assuming the wave is travelling at constant speed at this stage, then the layers B,C,D, ... etc

Figure 33. Structure of the knob.



between successive wave-fronts have equal volume. This procedure is not quite as artificial as might at first be supposed, because the estimates of size below show that layers B,C may be only cell thick. Nevertheless during the following description, let us assume that each layer behaves as a qualitative whole.

When the secondary wave hits a cell it becomes active for a period during which it actively submerges by amoeboid action into the interior of the tissue, by Hypothesis 2; it then ceases to be active and exudes slime. The active cells are in the shaded layers B and C in Figure 33. Therefore, going through the layers in detail we have :

A is static stalk, no longer active, with a cylindrical cellulose wall of exuded slime, shown by the thick line in Figure 33.

B is active, and the submerging cells of B cause B to minimise its surface area, at the same time as adhering to the top of A. Therefore if B has the volume of a sphere of diameter δ , then B will adopt a hemispherical top of radius δ , and cylindrical walls.

Lemma 6. The height of B, measured along this axis, is $\frac{2}{3}\delta$.

Proof. The volume of a sphere is $\frac{2}{3}$ that of an enclosing cylinder of equal height. Therefore a cylinder of equal volume has $\frac{2}{3}$ the height.

Estimates of size. The typical grex might contain 15,000 cells (it can vary from 500 to 50,000). About $\frac{2}{3}$ of these become spore cells, and half the rest comprise the foot, leaving 2,500 in the stalk itself. The stalk is about 1mm long (see the photograph in Figure 31(P)), and a cell is about 0.01mm in diameter [15], and so the stalk is about 100 cells long and 25 cells in cross-section. Therefore the stalk is about 5 cells wide on the average, and so δ is about 0.05mm, which agrees with the photographs in Figure 31. Therefore the volume of B is 5^3 times the volume of a cell, and consequently the number of active cells in B and C together is about 250.

Meanwhile the grex has ratio of length to width about 5:1, and allowing for the fact that cells in the grex are slightly elongated, the grex must be about 60 cells long and 16 cells wide. Therefore a cross-section of the grex contains about 256 cells. Therefore the number of cells active during the secondary wave is of the same order as the number of cells across a wave front of the primary wave. These

estimates help to show that our proposed knob structure is at least plausible. We now return to the business of analysing the layers of this structure.

- C is also active, and the cells have submerged by pulling the D layer over the top. C adheres to the surface of B. Since C has the same volume as B, and is at least one cell thick, and assuming B is 5 cells across, then C can only cover the hemispherical top of B (because a sphere of diameter 7 has roughly 3 times the volume of a sphere of diameter 5). Therefore the lower rim of C is at the point β in Figure 33.
- D is inactive because the secondary wave has not reached it yet. The cells of D are pulled radially inwards over the top of C, by the submerging cells of C, and are stretched taut in the process. The cells of D cannot slide over the cells of B and C, otherwise they would elastically slip back, exposing C again. Therefore the bottom rim of B cannot slide any higher than the point α in Figure 33, because this is the top-most point of the wall of slime. Therefore the layer D is stretched over the whole of the top of B and C, and, being the same volume, must be only half as thick as C. Therefore the cells of D must be stretched and elongated until they are only about half their normal width, as is confirmed histologically.

E, F, ... are as yet inactive, and are pulled up by D.

Therefore the outer cylinder of the knob consists of all the rest of the layers of stalk cells, which explains why the knob is cylindrical. Meanwhile the spore cells are less elastic than the stalk cells, and so at the presumptive stalk/spore frontier, which is the final wave-front, the knob broadens out into the shoulders of the fruiting body.

The above structure that we have proposed for the knob is somewhat complicated, and although we have indicated how the structure might have developed initially, the question remains whether the knob can preserve this structure. The answer is given by :

Lemma 7. The structure of the knob is time invariant.

Proof. We must verify that in one unit of time, each layer adopts the shape of its predecessor, translated a distance $\frac{2}{3}\delta$ upwards. We examine each layer separately :

A remains inactive and fixed.

B goes inactive and exudes slime. This enables the bottom rims of layers D,E,F to slide up to β , which is now the top of the newly made slime.

C remains active, and the effect of the submerging cells of C causes C to minimise its surface area, and round up to the same shape as B.

D becomes active as the secondary wave hits it, and submerges by pulling the cells of E radially inwards over the top, to form a closed layer over the top of D. Meanwhile the cells at the bottom rim of D submerge, by amoeboid action towards their neighbours in D, and therefore reduce the surface area of D by migrating from β up to γ .

E is not yet active, and is stretched over D with bottom rim held at β .

F, G, ... are pulled up.

Therefore the entire structure is reproduced isomorphically $\frac{2}{3}\delta$ higher.

Lemma 8. The diameter of stalk and rate of growth are proportional to the size of the fruiting body. Hence the time taken to culminate is independent of size.

Proof. Let T be the unit of time. Let A be the area of cross-section of the grex. Let V be the speed of the primary wave along the grex - we may assume for the moment that V is constant, since we are dealing with mid-growth stage. Then in time T the primary wave crosses volume VAT of the grex. This must equal the volume of one layer. But B has the volume of a sphere of radius δ . Therefore

$$\frac{4}{3} \pi \delta^3 = VAT,$$

$$\delta = \sqrt[3]{\frac{3VAT}{4\pi}}$$

Therefore the rate of growth of the stalk is

$$v = \frac{\frac{2}{3}\delta}{T} = \sqrt[3]{\frac{2VA}{9\pi T^2}}$$

Now T is a constant, because 2T is the time interval of the secondary effect, in other words the period that a cell spends actively submerging.

Suppose that we increase A by a factor λ , and the length of the grex by a factor μ . (Normally the ratio of the length to width of grex is 5:1, and so $\lambda = \mu^2$, but it is not necessary to assume this.) Then the length of travel of the primary wave is increased by μ , because of regulation, while the timing of beginning and end of the primary wave is unchanged, because it is determined by the development of tip and tail cells, independent of the size of grex. Therefore the speed V is increased by μ . Therefore VAT is increased by $\lambda\mu^2$, which is the increase in volume of the grex. Therefore δ , v are both increased by $\propto \sqrt{\lambda\mu^2}$, which is the increase in linear size of the fruiting body. This proves the lemma.

Remark. It might be possible to measure T microcalorimetrically. It is easy to measure A , and it might be possible to estimate V from detection of the primary wave. Hence one could predict and verify δ and v .

Lemma 9. The length of the knob shrinks at $\frac{1}{3}$ the speed that it grows upwards.

Proof. This lemma depends upon the experimental observation that, initially at any rate, the diameter of knob is twice the diameter of stalk [7]. Therefore area of cross-section of knob is 4 times that of stalk. Therefore the area of cross-section of the cylinder surrounding the stalk is 3 times that of stalk. But the upward growth is merely transference of pre-stalk material from cylinder to stalk. Hence the rate of growth of stalk is 3 times the rate of shrinkage of knob.

This concludes our discussion of structure of the knob and stage (d) of Figure 32; we now move on to the next stage.

(e) This stage occurs slightly after that shown in the photograph Figure 31(O), and the qualitative shape of the spore-mass is slightly more pronounced. We must explain this shape. The spore cells form an elastic mass, that, if it were allowed to float freely in a liquid of the same density, would minimise its surface area by adopting spherical shape. If this elastic mass is suspended in air from the top, then gravity would cause it to adopt a lemon-shape, as in Figures 31(P) and 32(g). At the stage we are considering in Figure 32(e) the spore mass is being pulled up from the top, but there is also a considerable frictional drag against the cylindrical stalk, and it is the combination of the top suspension together with downward drag, acting on the elastic

mass, that produces the characteristic shape.

The stalk is cylindrical, because this stage corresponds to a period during which the primary wave had constant velocity V , and so by Lemma 8 the stalk has constant diameter δ . The friction is increased by the fact that the slipperiness of the slime tends to get used up leaving a dry cellulose tube, so that the drag is more pronounced at the bottom. It should be possible to simulate these simple forces of elasticity, suspension, gravity and drag on a computer, and produce quantitative predictions about the qualitative shape of Figure 32(e).

(f) At this stage there occurs the characteristic hiccup*, as the spore mass suddenly flips from the dotted profile, similar to the previous stage (e), into the lemon-shape, similar to the secondary stage (g). The hiccup can be seen very clearly in the time-lapse film [7]. When the hiccup occurs the knob and shoulders do not move. Therefore the hiccup cannot be caused by a change in the upward supporting force, and must be caused by a sudden reduction of the downward frictional drag. And it is easy to see why, because the point x in Figure 32(f) is the point at which the stalk changes from being cylindrical to conical. Measurement of the stalk diameter at x and y , before and after the hiccup, [7], reveals that the diameter drops from about $\frac{1}{2}$ to about $\frac{1}{4}$ of the knob diameter. The change to conical shape at x causes the sudden disappearance of the maximal frictional drag of the cylinder on the bottom of the spore mass, and so friction suddenly becomes negligible compared with gravity in determining the shape of the spore mass. Hence the hiccup. Moreover it should be possible to simulate the hiccup by a minor adjustment of the previously suggested computer programme.

There remains the question of why the stalk diameter begins to decrease. In the proof of Lemma 3 we showed that the diameter depended upon V , A , and T . Now T is constant, being a property of the individual cell, and A is constant at this stage, since it is the cross-sectional area $\frac{1}{2}$ along the grex. Therefore our attention is drawn

* I am indebted to John Ashworth for pointing out the hiccup to me, after I had suggested that the morphogenesis might be caused by a secondary wave.

to V . And, sure enough, the primary wave slows down parabolically just before it stabilises by §9 Corollary 3.

This gives the correct qualitative explanation of why the diameter decreases. However we must be cautious about applying §9 Corollary 3 quantitatively, because in this application the cells do not stay topologically in the same place relative to one another during the morphogenesis. Therefore although the quantitative result is true for the primary wave, it may no longer apply to the secondary wave. Indeed, as the secondary wave comes to a halt, the continuity on which the estimate was based, becomes less important than the timing inside individual cells in determining their final position. This is an interesting and delicate point, and so let us enlarge upon it.

Suppose, on the contrary, that §9 Corollary 3 was applicable to the end of the secondary wave. Let V be the speed of the primary wave at time τ before its end, and let δ be the diameter, v the speed, and η the distance of the stalk below its eventual top, at time τ before the end of the secondary wave. Then $V \sim \tau$, by §9 Corollary 3, and so

$$\delta \sim v \sim V^{1/3} \sim \tau^{1/3}$$

by Lemma 8 above, implying

$$\eta \sim \int v d\tau \sim \tau^{4/3} \sim \delta^4 .$$

Therefore this would predict a very blunt 4th power top to the stalk. Now the initial narrowing of the stalk, when it changes from cylindrical to conical shape at the point x in Figure 32(f), may indeed obey a 4th power law, and it would be interesting to measure this experimentally. However when the stalk has reduced to a width of 1 or 2 cells then this result no longer applies because of the individuality of cells, as follows.

The final wave front of the primary wave contains about 250 cells. Therefore we may expect that the last few of these cells will experience their catastrophic switch from pre-spore to pre-stalk not all together, but spaced out, at individual times. How few, and how spaced-out, will depend upon the noise level (for instance the amount of irregularity in cell size and original environment). Therefore as the secondary wave hits each of these last few cells in succession, it will induce it to slide

up and adhere to the top of its predecessor. Therefore the final piece of stalk will be only one cell thick, and possibly of length comparable to the length of the final spore mass (which is about 30 cells). This can be seen in the final enlarged shot of the film [7], for instance, where the stalk is only 1-cell wide at the base of the spore mass.

(g) As the last stalk cells submerge, they pull spore cells over the top of the knob. Therefore in its final stages the knob consists of spore cells, rather than stalk cells. Therefore it is rounded rather than cylindrical. The final disappearance of the knob is due to the elasticity of the spore mass trying to reduce its surface area, which is a weaker force than that created by cells submerging. Added to this is a renewed frictional drag, because the top part of the stalk is a cylinder again, one cell wide. Therefore the final disappearance of knob takes a long time, as in the film [7], and it may not entirely disappear. This effect will be exaggerated in fruiting bodies originating from an irregular environment. This completes our description of the seven stages in Figure 32.

17. EXPERIMENTS TO BE DONE.

Summarising, we have explained the morphogenesis of the culmination of cellular slime mold, assuming only :

- the shape before culmination,
- the differentiation between spore and stalk,
- that differentiation begins at the tip of the grex, and
- that there is a secondary wave of cells submerging and exuding slime.

Our theory agrees with the classical experimental observations of Bonner [2, and 21 p. 4363] and with many qualitative details found by Gerisch [7] and Farnsworth [5]. The new contribution of our theory, compared with previous theories, is that it offers an explanation of why the local cellular forces occur in the order that they do, and how they create the surprisingly complicated sequence of global shapes of the culminating fruiting body. In other words it provides a link between the local and the global in space-time.

The theory admits several testable predictions about timing and the structure of the knob, and some of the experiments that could be done.

(i) Establish the timing and speed of the primary wave by cutting different proportions off the front of the migrating grex at different times. It might be possible to first slice the grex lengthways, in order to be able to use half as control.

(ii) Verify and time the secondary wave, by grafting pieces of pre-stalk onto pre-spore and observing when the cells submerge.

(iii) Measure the duration of the secondary effect microcalormetrically.

(iv) Verify the diameter and growth speed given by Lemma 3.

(v) Verify the knob shrinkage speed given by Lemma 4.

(vi) Simulate the spore mass shape and hiccup on a computer.

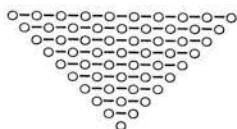
(vii) Turn the culminating fruiting body upside down. Then after the hiccup, when gravity becomes dominant, the fruiting body should be fatter because it is suspended from below.

(viii) Cut off half the knob during culmination. Then the spore mass and the remaining cut cylinder of knob stump should slide down a little, leaving an exposed stump of stalk. Then when the secondary wave hits the knob stump it should submerge and form a submerged solid torus of stalk cells. Similarly one should be able to design several tricks to test the knob structure, by making the secondary wave submerge in various different directions, (as for example in [5]).

18.

CONCLUSION

The point of view in this paper, and the explanations, designs of experiments and predictions in the two examples, arise from the use of catastrophe theory. If primary and secondary waves are confirmed to be as widespread as it seems they must be, then this by itself would be a useful contribution of catastrophe theory to biology. But we have only used the two simplest elementary catastrophes, the fold and the cusp. To glimpse the potential riches offered by the use of the higher dimensional catastrophes, one only has to glance the astonishing writings [18,19] of the creator of this theory, René Thom.



REFERENCES

1. B.I. Balinsky, An introduction to embryology, (Saunders, Philadelphia & London, 1965) 2nd edition.
2. J.T. Bonner, The cellular slime moulds, (Princeton Univ. Press, 1967).
3. J. Cooke, conversations.
4. J. Cooke, Some current theories of the emergence and regulation of spatial organisation in early animal development, Annual Rev. of biophys. and bioeng. (1975), to appear.
5. P. Farnsworth, Morphogenesis in the cellular slime mould Dictyostelium discoideum; the formation and regulation of aggregate tips and the specification of developmental axes, J. Embryol. exp. Morph. 29 (1973), 253-266.
6. D. Garrod & J.M. Ashworth, Development of cellular slime mould, Dictyostelium discoideum, Symp. Soc. Gen. Microbiology XXIII, Microbial differentiation (1973), 407-435.
7. G. Gerisch, Dictyostelium discoideum (Acrasina) Aggregation und Bildung des Sporophors (Institut für den Wissenschaftlichen Film, Göttingen), Film E 631 (1963).
8. T. Gustafson & L. Wolpert, The forces that shape the embryo, Discovery 22 (1961), 470-477.
9. J. Koyama, Normal table of the development of the Japanese newt Diemictylus (Triturus), Zoo. Mag. (Tokyo) 42 (1930), 465-473.
10. C.K. Leach, J.M. Ashworth & D.R. Garrod, Cell sorting out during the differentiation of mixtures of metabolically distinct populations of Dictyostelium discoideum, J. Embryol. exp. Morph. 29, 3 (1973), 647-661.
11. W. Luther, Entwicklung des Molcheies (Institut für den Wissenschaftlichen Film, Göttingen) Film C. 939 (1967).
12. P.D. Nieuwkoop, conversations.
13. A.W. Pollister & J.A. Moore, Tables for the normal development of Rana sylvatica, Anat. Rec. 68 (1937) 489-496.

14. K.B. Raper, Pseudoplasmodium formation and organisation in *Dictyostelium discoideum*, Jour. Elisha Mitchell Scientific Society, 56, (1940), 241-282.
15. A. Robertson, Quantitative analysis of the development of cellular slime mold, (A.A.A.S., 1970, Some Mathematical Questions in Biology, V,), Lectures on Maths. in the Life Sciences, Vol. 4 (Amer. Math. Soc., Providence, U.S.A., 1972), 47-73.
16. S.N. Sedra & M.I. Michael, Normal table of the Egyptian Toad, *Bufo regularis* Reuss, Československá Morfologie, 9 (1961), 333-351.
17. W. Shumway, Stages in the normal development of *Rana pipiens* I, Anat. Rec. 78 (1940), 137-147; II, Anat. Rec. 83 (1942), 309-315.
18. R. Thom, Stabilité structurelle et morphogénèse, (Benjamin, New York, 1972).
19. R. Thom, A global dynamical scheme for vertebrate embryology, (A.A.A.S. 1971, Some Mathematical Questions in Biology, VI), Lectures on Maths. in the Life Sciences, Vol. 5 (Amer. Math. Soc., Providence, U.S.A., 1973) 3-45.
20. W. Vogt, Gestaltungsanalyse am Amphibienkeim mit örtlicher Vitalfärbung, II, Gastrulation und Mesodermbildung bei Urodelen und Anuren, Roux Arch. 120 (1929), 385-706.
21. C.H. Waddington, Principles of embryology, (Allen & Unwin, London, 1956).
22. C.H. Waddington, The strategy of the Genes, (Allen & Unwin, New York, 1957).
23. P. Weiss, Principles of development, (1939; reprinted Hafner, New York, 1969).
24. E.C. Zeeman, Differential equations for the heartbeat and nerve impulse, Towards a theoretical biology 4, (Ed. C.H. Waddington, Edinburgh University Press, 1972), 8-67.
25. E.C. Zeeman, Differentiation and pattern-formation, (Appendix to Reference 4).
26. E.C. Zeeman, Gradients and catastrophes in developmental biology, (in preparation).