Sev CN CT

8. THE THEORY OF THÜNEN SYSTEMS

8.1. Introduction

A

~~One often finds,~~ both in nature and society, some spatial patterns ~~which~~ may be described roughly as follows. There is a certain special location surrounded by a series of concentric "rings" or "shells". At the locations in any one ring the same activity is occurring. From ring to ring there is a tendency for activities to become less "intense" in some sense as one moves outward.

10 x 27
12 TR
φ = 2 ems

An example is provided by a sphere in gravitational equilibrium. Here the densest substances lie toward the center and density declines as one moves outward, ending in an ever thinner atmosphere. The environs of a volcano provide a less clear-cut example, as does an organic cell with its nucleus and cytoplasm.

In the social world ~~one has the~~ a pattern of agricultural land uses surrounds a city. These tend to decline in "intensity" with increasing distance from the city. Within the city itself ~~one has~~ are the highly intensive land uses of the central business district (CBD) and a gradual diminution of intensity as one moves outward. On the "micro" level, the fields of individual farms tend to be cultivated with diminishing intensity as one moves farther from the farmhouse. We might also refer to the distribution of onlookers at sports events and other spectacles.

We ~~shall~~ discuss these and other examples in greater detail later. They are adduced here merely to introduce the class of phenomena to be considered. ~~Note that~~ in all cases there is a greater or lesser degree of distortion from the "ideal" pattern of concentric rings of homogeneous activities. ~~Note also that~~ the same general phenomenon can occur at very different scales of magnitude: from the spatial ordering of an individual household to the pattern of large geographic regions, and even — (as we shall see) — to the entire world economy.

(w) These patterns will be called Thünen systems, after the man who first investigated one of them in society, ~~namely,~~ viz. the pattern of agricultural land uses around a city in an isolated region.[1] We ~~shall~~ make no use of Thünen's specific formulations, however, because modern developments have corrected and generalized them considerably, and ~~the present~~ this chapter will generalize them even further.

In developing a theoretical model for Thünen systems, the first problem that arises is ~~that of~~ specifying precisely what ~~one means~~ is meant by such systems. The concepts "activity" and "intensity", even the concept of "concentric rings", were used above in a vague commonsense way and need explication. After doing this, we present a model ~~which~~ that is both explanatory and optimizing; — ~~that is~~ i.e., it shows how the Thünen pattern will arise from the behavior of individual agents, and also demonstrates that this pattern solves an optimization problem.

We ~~shall~~ place greatest stress on this optimality feature of Thünen systems, because it has been largely neglected in the past. ~~In fact,~~ it is shown that Thünen systems are optimal for a special case of the measure-theoretic transportation problem of chapter 7, ~~called~~ the allotment-assignment problem. (Certain special Thünen systems also optimize the allocation-of-effort problem discussed in chapter 5). The potentials for this transportation problem may be interpreted (in part) as land values, and this establishes the connection between the optimizing and explanatory aspects of the model.

The special location ~~which~~ that is the center of symmetry of the Thünen rings will be called the nucleus (corresponding in the examples above to the city, the CBD, the farmhouse, etc.). This will play a basic rôle for most of our exposition. In the end, however, ~~it will turn out that~~ even the nucleus can be dispensed with. The essential point is that the "desirability" of a location can be summarized in a single real number. This is usually the "distance" from the nucleus, but may be well-defined even if there is no nucleus. All these points will be elaborated on below.

g25-2

A

8.2. Ideal Distances and Ideal Weights

We now re-introduce our three basic sets: Resources, Space and Time (R,S,T). ~~Actually,~~ the formal model to follow makes no concrete assumptions about the nature of these sets, and the generality resulting from this fact is useful. For example, T can be interpreted as having a bounded horizon, or as being discrete, rather than as being the whole real Time-axis. S may be thought of as a limited region of the Earth's surface, rather than as all of Space. Similarly, R may be thought of as restricted to ~~those~~ resource-types ~~which~~ that make sense in the Thünen context — e.g. those ~~which~~ that are "transportable".

We suppose ~~that there is~~ a real-valued function θ with domain $R \times T \times S \times S$ exists, the unit transport cost function. Specifically, $\theta(r,t,s_1,s_2)$ is the cost of shipping unit mass of resource type r at time t from origin s_1 to destination s_2.

D

Definition: $\theta: R \times T \times S \times S \to$ reals is factorable iff ~~there are exist~~ two functions, $g: R \times T \to$ reals and $h: S \times S \to$ reals, exist such that

$$\theta(r,t,s_1,s_2) = g(r,t)\, h(s_1,s_2), \qquad (8.2.1)\ \cancel{(1)}$$

for all $r \in R$, $t \in T$, $s_1, s_2 \in S$. Function $g(r,t)$ is called the ideal (or economic) weight of resource r at time t, and $h(s_1,s_2)$ is called the ideal (or economic) distance from s_1 to s_2.

Excluding the trivial case when θ is identically zero, ~~one easily~~ we can establish~~es~~ that g and h are unique up to scalar multiplication. To be precise, if g and h satisfy (1), ~~then~~ so does the pair gx, h/x — (x being any non-zero real number) — and these pairs are the only solutions. Also, if θ is non-negative, and there exist g, h satisfying (1), then ~~there exist~~ non-negative g, h satisfying (1) exist. For the following discussion we assume g and h are non-negative.

Consider the economic meaning of θ and the condition (1). There is, first ~~of all~~, the problem of what instant "t" refers to in the case of time-consuming trips. A simple convention takes t to be the average of time of departure from s_1 and time of arrival at s_2. (A more elaborate analysis would insert an extra time component, resource r departing from s_1 at t_1 and arriving at s_2 at t_2. But this elaboration is not needed for the problems of this chapter).

The mass flowing through the transportation system will be represented by a measure μ on universe set $R \times T \times S \times S$: $\mu(E \times F \times G \times H)$ = total mass of resources of types E flowing at times F from sources in region G to sinks in region H. Given μ and θ, total transport cost incurred is assumed to be

$$\int_{R \times T \times S \times S} \theta \, d\mu \quad . \qquad (8.2.2)\ \cancel{(2)}$$

This is a severe assumption, ignoring as it does large-lot economies in transportation, congestion and other interaction effects. A few devices mentioned below help to overcome these limitations, but are only partially successful.

~~As for~~ the factorability condition (1), ~~it~~ states in effect that no source-sink pair (s_1, s_2) has a comparative advantage over any other such pair for the shipment of any resource at any time vis-à-vis another resource at another time. This is again a strong condition, and it is easy to find situations where it breaks down. For example, let s_1 and s_2

have good pipeline connections and poor road connections, and *vice versa* for s_3 and s_4. Then s_1 and s_2 might be "closer" for oil transportation, and "further apart" for passenger transportation, than s_3 and s_4; θ is clearly not factorable in this case. Nonetheless, we ~~shall~~ assume factorability as a very useful first approximation.

The great simplification that arises from factorability is that the same spatial transport-cost pattern applies to all resource-types and times, and may be summarized in a single function having only spatial arguments, ~~namely~~ viz. the ideal distance function h.

g25-3. ~~Let us~~ We now examine the two ideal functions, g and h, which arise from a plausible fac~~o~~torable transport-cost function θ. As noted ~~above~~, g and h are unique up to a scalar multiple, so that the ratios of non-zero values $g(r_1,t_1)/g(r_2,t_2)$ are 1 1/8
uniquely determined by θ, and similarly for h. The resulting patterns need not have any close relation to *physical* weights or distances, respectively, though there will presumably be some overall positive correlation between ideal and physical values.

Consider the weight function, g. Resource-types which, for given physical weight, are bulky, ~~or~~ valuable, ~~or~~ heavily taxed, or need special handling, will tend to have relatively high ideal weights. ~~Also,~~ Small-lot shipments of the same resource tend to cost more per unit weight than large-lot shipments. It seems at first that the linearity of (2) precludes taking account of this last phenomenon, but one possible device for doing so is to distinguish different-size packages of the same resource formally as distinct resource-types, the larger packages having smaller ideal/physical weight ratios.

How will $g(r,t)$ vary with time, for fixed r? The secular trend will usually be downward, for two reasons. First, there are technological improvements in transport and communications, extensions of the various grids, more vehicles in existence, etc., all of which reduce real transportation costs.[2] The second reason is the need to *discount*. To make the cost contributions of different times comparable in the integral (2), they must all be discounted to the same moment. The easiest solution ~~way to do this~~ is to build the discount factor directly into the ideal weight function g. The same real cost in the far future is less weighty than in the near future, and discounting introduces an additional "levitational" force over time.

Congestion may sometimes be allowed for by adjusting ideal weight. For example, suppose one is studying a metropolitan area, and ~~that~~ congestion appears periodically at weekday rush hours. One can represent this by letting g rise at these times; Things get "heavier" during rush hours. (This is another device for circumventing in part the restricted form of

(2). An adequate theory of congestion would require total cost to be a non-linear function of μ, however.

Factorability implies that, for any particular resource-type *r*, unit transport costs rise or fall proportionally for all source-sink pairs over time. Thus transport innovation must reduce costs *pro rata*; a reduction in one region but not in another would violate factorability. Similarly, congestion must raise costs proportionally on all routes. These unlikely circumstances underline the strength of the factorability assumption.

Turning to ideal distances, we note ~~to begin with~~ that the term "distance" is a misnomer, because *h* need not obey the metric postulates. In particular, the symmetry postulate may be violated, due to up-hill vs. down-hill movements, wind and water currents, one-way streets, tariffs on imports but not exports, etc. Ideal distances will be distorted from physical distances because of geographic irregularities, because some pairs of locations have "good connections" relative to others, because fares are not faithful reflections of distances, because of heavy taxation at border crossings, etc. Just as temporal variations of congestion can be allowed for by adjusting ideal weights, spatial variations can be allowed for by adjusting ideal distances. That is, if certain regions — (such as the central portions of cities) — are generally congested, ideal distances between points in these regions will be large relative to physical distances. Speaking broadly, ideal distance tends to increase less than in proportion to physical distance (except perhaps for very long trips) because ~~. The main reason is that~~ "over-head costs" such as loading, packing, billing, getting up steam, etc. (which may be a substantial fraction of total transport costs) — are spread over a larger physical distance. (For very long trips the factors of cumulative fatigue and spoilage, and the need to carry large amounts of food and fuel, work in the opposite direction. For rocket flights, ~~the longest trips of all~~, the fuel carrying ~~carriage~~ factor is crucial).[3]

8.3. Ideal Distances in Thünen Systems

We ~~shall~~ develop several variant models for Thünen systems. The one to which most attention will be devoted is the entrepôt model. In this section we ~~shall~~ concentrate on some of its formal characteristics, and do not worry about its realism.

The distinguishing feature of entrepôt models is the existence of a special location, called the nucleus, having the property that all transportation flows have the nucleus either as origin or as destination. That is, the exports of any land use located anywhere in the system all go to the nucleus; the imports of that land use all come from the nucleus. The nucleus functions as an entrepôt in the sense that a shipment from location s_1 to s_2 can be accomplished in two steps: from s_1 to the nucleus, and from the nucleus to s_2. "Foreign trade" (i.e. ~~that is,~~ flows between locations in the Thünen system and locations outside it — is not excluded, but any such trade must be channeled through the nucleus, so that the nucleus also functions as a gateway between the system and the rest of the world.

For entrepôt models we postulate a transport cost function with a slightly modified factorability condition. ~~First of all,~~ transport cost between two non-nuclear sites is irrelevant, since by assumption no such flows ever occur. Hence we need to postulate the factorability condition, (2.1) ~~(1) of section 2,~~ only in the case where s_1 or s_2 is the nucleus. Formally, the unit transport cost function θ satisfies the following condition:

~~There exist~~ two functions $g_{in}, g_{out}: R \times T \to$ reals, and two functions $h_{in}, h_{out}: S \to$ reals, exist such that

$$\theta(r,t,s,s_N) = g_{in}(r,t)h_{in}(s), \qquad (8.3.1)\ (1)$$

~~and~~

$$\theta(r,t,s_N,s) = g_{out}(r,t)h_{out}(s). \qquad (8.3.2)\ (2)$$

Here s_N denotes the nucleus. Comparing (1) ~~here~~ with (2.1) ~~(1) of section 2,~~ we see that $h_{in}(s) = h(s,s_N)$. The second argument of h is fixed at s_N, and is dropped for simplicity. Thus $h_{in}(s)$ is simply the ideal distance from location s to the nucleus. Similarly, $h_{out}(s)$ is the ideal distance to location s from the nucleus.

The two g functions have a different significance. We are allowing the same resource r at the same time t to have two different ideal weights, depending on whether it is flowing into ~~the nucleus~~ or out of the nucleus. This is a further relaxation of the factorability condition (2.1) ~~(1) of section 2.~~

Conditions (1) and (2) together are clearly weaker than (2.1) ~~(1) of section 2.~~ In fact they are a bit too weak for our purposes, and we now add the symmetry condition that

$$h_{out} = h_{in}. \qquad (8.3.3)\ \cancel{(3)}$$

That is, the ideal distance between the nucleus and any other location in the system is the same in both directions. We denote this common function by h. ~~(Note that~~ the domain of h is now simply S, not $S \times S$ as it was in section 8 2).

Function $h(s)$ will be referred to simply as "the distance of s". It provides a general index of inaccessibility of locations in the entrepôt model. The fact that the relative locational advantages of different places can be summarized in a single number in this way is one essential precondition for the striking simplicity of the results obtained for Thünen systems. (ii)

An example or two will illustrate that the symmetry assumption is less restrictive than might appear at first glance. Suppose the nucleus is perched on top of a hill, so that it costs, say, twice as much to transport resource r at time t from location s to the nucleus as to go in the opposite direction. Then, if (1) and (2) are satisfied, so is the symmetry condition. We merely take $g_{in} = 2 \cdot g_{out}$, $h_{in} = h_{out}$, so that a resource is twice as "heavy" traveling to the nucleus as when traveling away from it. ~~Note that~~ we have merely thrown the burden of representing cost differentials onto the weight function, leaving the distance function invariant, just as we did in the case of the "shrinking globe".

The situation opposite to the one just mentioned is probably more common in practice: All roads lead to Rome more readily than they lead away ~~from Rome~~, because of asymmetries of information. In this case g_{in} is smaller than g_{out}, and the same argument applies.

9.25-5 ~~Again,~~ consider the rush hour phenomenon in big cities. In the morning it is easier to travel away from the central business district than toward it, and the reverse is true in the evening. g_{in} would then be larger than g_{out} at morning times and smaller at evening times, and the symmetry condition (3) would not necessarily be violated.

We ~~Let us~~ now suppose that conditions (1) & (3) ~~(1), (2) and (3)~~ obtain, so that the distance function $h: S \to$ reals is well-defined. The region

$$\{s \mid h(s) < z\} \tag{4}$$

(8.3.4)

is then called the open disc of radius z (about the nucleus). (A similar concept has already been defined for metrics, but we may not be dealing with a genuine metric in this case. Still, the concept is well-defined if h is).

The shape of the region (4) will ~~of course~~ depend on the nature of the function h. Suppose ~~that Space~~ S is the plane, and for convenience let the nucleus s_N be at the origin, (0,0). If h is derived from a Euclidean metric, ~~then~~ the regions (4) will be circular discs centered on the origin. If $h(s) = |x| + |y|$ (where (x,y) are the cartesian coordinates of s), then the regions (4) will be diamonds, i.e. ~~— that is~~, squares with sides at 45° to the X- and Y-axes. This arises from a city-block metric, which in turn may be thought of as arising from a road system permitting only motions parallel to the X- or Y-axis.

cap X, Y

Another common case arises from a limited number of traffic arteries converging radially on the nucleus (road, rail, river). ~~etc.~~ Travel is relatively easy along such radials and difficult off them. In this case the regions (4) will tend to be amoeboid-shaped, with "pseudopods" projecting out along each artery. It is even possible for these regions to fall into several disconnected pieces. This occurs with limited-access transportation systems (highways with infrequent exits, railways, airports ~~, etc.~~). Here the immediate neighborhood of a point of access to the transportation network may be an isolated outpost that ~~which~~ is economically "close" to the nucleus though physically distant.

The significance of this discussion is that in Thünen systems land uses are arranged in "rings," which are set-theoretic differences of open discs (4) with different radii.

Only in the Euclidean case will these literally be rings, ~~that is,~~ i.e., annuli centered on the nucleus. In other cases these rings will be more or less irregular and even disconnected. We would expect, for example, that typically "urban" land uses would "sprawl" deep into the countryside along major radial arteries, and ~~that they~~ would tend to appear in the vicinity of commuter railway stations and airports.

These diverse phenomena are all covered by the entrepôt model, which predicts the pattern of land uses in terms of <u>ideal</u> distances. The geographical implications will then depend on the shape of the regions (4). The model itself, however, does not need ~~and does not~~ nor make any such assumptions, but is formulated throughout in terms of ideal ~~distances~~, not physical distances.

8.4. <u>Land Uses</u>

The <u>spatial</u> <u>field</u>, with the particular structure just discussed, is one of the two basic ingredients ~~which~~ constitute ing Thünen systems. The other is the set of <u>land</u> <u>uses</u> ~~which are~~ to be distributed over this field. We now discuss these: first, more or less formally, and then with concrete interpretations and illustrations of the concepts involved.

Formal Structure

We ~~shall~~ use a version of the activity analysis model of ~~chapter 4,~~ section 4.5 and ~~Let us briefly~~ review the salient concepts, using the notation of that section. An activity, q, is a triple of measures, ρ on (Ω_r, Σ'), and λ_1 and λ_2 on $(R \times T, \Sigma_r \times \Sigma_t)$. Here Ω_r is the space of transmutation-paths, and ρ represents the "capital-goods" structure of the activity; λ_1 is the "production" measure, describing the resource-time distribution of outputs; similarly, λ_2 is the "consumption" measure.

This describes one activity. Next, Q is the set of all feasible activities, and ν, a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$, describes the assignment of activities to locations. On measurable rectangles, $\nu(E \times F)$ is the "amount" of activities of types F located in region E. This determines the total production measure μ_1 over the space $(R \times S \times T, \Sigma_r \times \Sigma_s \times \Sigma_t)$ as follows:

12 pt brackets

$$\mu_1(G) = \int_{S\times Q} \lambda_1\{q, \{(r,t) \mid (r,s,t) \in G\}\}\nu(ds, dq), \qquad (8.4.1)\ ~~(1)~~$$

for all $G \in \Sigma_r \times \Sigma_s \times \Sigma_t$. Here λ_1 is the function with domain $Q \times (\Sigma_r \times \Sigma_t)$ for which $\lambda_1(q,\cdot)$ is the production measure associated with activity $q \in Q$. The λ_1 is assumed to be an abcont conditional measure, which insures that the integral (1) is well-defined, and that μ_1 is a measure. Similarly, conditional measure λ_2 is constructed from the consumption measures of the various activities. Replacing λ_1 by λ_2 in (1), we obtain μ_2, the total consumption measure over $R \times S \times T$ determined by ν.

~~Let us~~ We now place these concepts in the Thünen context. The essential point is that all production (no matter when, where, or what is produced) must get shipped to the nucleus. Similarly, all consumption (over all Time, Space, and ~~Resources~~) must come from the nucleus. Combined with our factorability assumptions, this yields an expression for the total transport cost incurred by an assignment ν. Furthermore, we are able to apply the concept of ideal weight to the activities themselves, not just to resource-time pairs; this simplifies things considerably.

We now spell out these statements. ~~As discussed above, there is~~ an ideal distance function $h{:}S \to$ reals exists giving the "inaccessibility" of any location from the nucleus, and two ideal weight functions $g_{in}, g_{out}{:}R \times T \to$ reals exist. We assume these functions are nonnegative and measurable. Define the in-weight of activity q as follows:

$$w_{in}(q) = \int_{R\times T} g_{in}(r,t)\,\lambda_1(q,dr,dt). \qquad (8.4.2)$$

The out-weight of activity q, written $w_{out}(q)$, is defined by (2) by substituting g_{out} for g_{in}, and λ_2 for λ_1. Finally, the weight of activity q is the sum of these two:

$$w(q) = w_{in}(q) + w_{out}(q). \qquad (8.4.3)$$

The w_{in}, w_{out} and w are all extended real-valued functions with domain Q. The conditions on g_{in}, g_{out} and λ_1, λ_2 insure that they are non-negative and measurable.

We now show that the total transport cost incurred in a Thünen system under assignment measure ν is simply

$$\int_{S\times Q} h(s)w(q)\nu(ds,dq). \qquad (8.4.4)\ ~~(4)~~$$

~~First of all,~~ total transport cost is the sum of cost incurred on shipments *into* the nucleus plus cost incurred on shipments *out of* the nucleus. The in-shipment cost is given by

$$\int_{R\times S\times T} g_{in}(r,t)h(s)\,\mu_1(dr,ds,dt), \qquad (8.4.5)\ ~~(5)~~$$

since this is what (2.2) ~~(2) of section 7~~ reduces to for the special case in hand. Here μ_1 is the total production measure as given by (1) ~~above~~. (Remember that all the mass of the distribution μ_1 must be shipped to the nucleus. A unit mass of resource type *r* located at *s* and shipped at moment *t* incurs a cost of $g_{in}(r,t)h(s)$.)

We claim that (5) is equal to

$$\int_{S\times Q} h(s)w_{in}(q)\nu(ds,dq). \qquad (8.4.6)\ ~~(6)~~$$

925-7

To show this, we introduce the measure $\mu_1{}^*$ on the product space $S \times Q \times R \times T$ by the following iterated integral:

$$\mu_1{}^*(H) = \int_{S\times Q} \nu(ds,dq) \int_{R\times T} \lambda_1(q,dr,dt)I_H, \qquad (8.4.7)\ ~~(7)~~$$

for all $H \in \Sigma_s \times \Sigma_q \times \Sigma_r \times \Sigma_t$. (Here I_H is the indicator function.) ~~One verifies~~ We verify that μ_1 given by (1) is the marginal of $\mu_1{}^*$ on the component space $R \times S \times T$.[4]

It follows from the induced integrals theorem that (5) is equal to

$$\int_{S\times Q\times R\times T} g_{in}(r,t)h(s)\,\mu_1{}^*(ds,dq,dr,dt).$$

By (7) and Fubini's theorem, this in turn equals the iterated integral

$$\int_{S\times Q} \nu(ds,dq) \int_{R\times T} \lambda_1(q,dr,dt)\,g_{in}(r,t)h(s). \qquad \cancel{(8)}\ (8.4.8)$$

Evaluating (8) from right to left, the integration over $R \times T$ yields the simple expression $w_{in}(q)h(s)$, by (2), so that (8) equals (6). We have proved that (5) and (6) are indeed equal.

The out-shipment cost is given by (5), with g_{out} and μ_2 replacing g_{in} and μ_1, respectively. The argument just given, with λ_2, $\mu_2{}^*$ replacing λ_1, $\mu_1{}^*$, proves that the out-shipment cost is equal to

$$\int_{S\times Q} h(s)w_{out}(q)\,\nu(ds,dq). \qquad \cancel{(9)}\ (8.4.9)$$

Finally, adding (6) and (9), and using (3), we see that total transport cost is indeed given by (4). This completes the proof.

18 In common sense terms, the argument just given amounts to the following. An activity determines a certain production and consumption pattern of resources over time. These incur transport costs per unit ideal distance as determined by their ideal weights, and this implicitly determines a weight for the activity itself, ~~namely,~~ viz. the cost incurred by its inputs and outputs in moving unit distance. This activity weight is given by (2) and (3). It is then intuitively plausible that the total transport cost incurred by the spatial activity distribution ν should be given by (4), the integral of the activity weights multiplied by the ideal distances its inputs and outputs must travel.

~~There~~ are important advantages obtained by this transformation. ~~First,~~ the expression (4) in terms of activities is much simpler than ~~the expression~~ (5) plus the corresponding expression for out-shipments. ~~In fact,~~ using (4) and the other constructions discussed below, it is possible to dispense with explicit consideration of ~~Resources~~ and ~~Time~~, and to work entirely with activities (and ~~Space~~). This is the natural approach ~~when one comes to~~ for concrete applications and it again leads to great formal simplicity.

Note that the "capital-goods" structure of activities, given by the measure space $(\Omega_r, \Sigma', \rho)$ (where ρ depends on activity q), does not influence transportation cost. This is as it should be; ρ refers to the "internal" operation of these activities, and no spatial movement is involved. (Shipments of equipment, construction materials, etc., are already incorporated in λ_1 and λ_2.) ρ, in fact, plays a very subordinate rôle in what follows, and will be ignored except for occasional comments.

We now come to the question of constraints on the possible activity distributions ν. Just one kind of constraint will be imposed: a limit on areal capacity.[5] That is, activities demand "room" in which to operate; regions have a limited amount of "room", and this limits the total amount of activities ~~which~~ that can be squeezed into them.

In ~~chapter 4,~~ section 4.5, the areal capacity constraint was written in the following form:

$$\int_{F\times Q} k \, d\nu \le \alpha(F), \qquad \overset{(8.4.10)}{\cancel{(10)}}$$

25-8

for all regions F. Here α is a measure on physical Space, (S,Σ_S), the <u>ideal</u> <u>areal</u> measure. The non-negative measurable function $k: S \times Q \to$ reals gives the "demand for room" by activity q at location s. (10) then states that the total demand for room in region F cannot exceed the capacity of that region.

We ~~shall~~ make the special assumption that $k = 1$ identically. (10) then becomes

$$\nu(F \times Q) \le \alpha(F), \qquad \overset{(8.4.11)}{(11)}$$

for all regions F.

The step from (10) to (11) is less restrictive than it appears to be. It amounts, essentially, to the assumption that function k is factorable: $k(s,q) = k_1(s)k_2(q)$ for some pair of positive functions $k_1: S \to$ reals and $k_2: Q \to$ reals. To see this, define two new measures, α' on S, and ν' on $S \times Q$ as follows:

$$\alpha'(F) = \int_F \frac{1}{k_1(s)} \alpha(ds)$$

$\int_F [1/k_1(s)] \alpha(ds)$

for all $F \in \Sigma_s$, and

$$\nu'(G) = \int_G k_2(q) \nu(ds,dq) \qquad \text{(8.4.12)} \; \cancel{(12)}$$

for all $G \in \Sigma_s \times \Sigma_q$. Then from (10) we obtain for all $F \in \Sigma_s$,

$$\int_{F \times Q} k_1(s) \nu'(ds,dq) = \int_{F \times Q} k_1(s) k_2(q) \nu(ds,dq) \le \alpha(F) = \int_F k_1(s) \alpha'(ds).$$

on all on one line

Treating the left- and right-~~hand~~ integrals as measures over S, we integrate the positive function $1/k_1$ with respect to them to obtain

$$\nu'(F \times Q) \le \alpha'(F), \qquad \text{(8.4.13)} \; \cancel{(13)}$$

for all $F \in \Sigma_s$. Condition (13) has the same form as (11).

Now the units in which activities are measured are arbitrary, and the "amount" of activity has no intrinsic meaning. Suppose, then, we change measurement units as follows. Activity q (or, more precisely, unit level of activity q) is now redefined to be the triple

48.

$$\left(\frac{\rho}{k_2(q)}, \frac{\lambda_1}{k_2(q)}, \frac{\lambda_2}{k_2(q)}\right),$$

where $(\rho, \lambda_1, \lambda_2)$ is the original activity q. Then assignment ν in the original units is the same as ν' in the new units, ν and ν' being related by (12). Similarly, there is no intrinsic significance to the ideal areal measure α, and it might just as well be replaced by α', with corresponding changes in k to keep the constraint conditions invariant.

With these changes of units, (10) becomes (13). We may, in fact, simply forget about the original measures ν and α, and drop the primes in (13), obtaining (11). (Corresponding changes must also be made in the weight function $w(q)$; we suppose this has been done, without changing notation).

One can now give an intuitively appealing interpretation to the nondescript concept of "amount" of activities, ν. Constraint (11) implies that ν and α are dimensionally comparable, so that ν may be thought of as given in "ideal" areal units ("acres", if you wish). Specifically, $\nu(F \times G)$ is the "acreage" required by the activities of types G, which are operating in region F. Similarly, the measures ρ, λ_1, λ_2 have the dimensions "mass per unit area". For example, $\lambda_1(E \times H)$ would be the production of resources of types E in period H, in "tons per acre", say.

We have been discussing "activities" in general ~~up~~ to this point. Let us refer to activities ~~which~~ that have a <u>positive</u> demand for "room" as <u>land uses</u>. All the activities discussed in connection with Thünen systems will ~~in fact~~ be land uses, as is clear from constraint (11). (Non-land-using activities could not be measured in acres ~~of course,~~ since positive amounts ~~of them~~ could be operating in regions of zero ideal area).

The areal constraint (11) is expressed as an inequality. In what follows we ~~shall~~ find it convenient to express this as an equality. No real loss of generality is involved here, since we can add a special land use called "vacancy" which takes up the slack, if any.

This concludes our formal discussion of land uses. The two basic formulas we have arrived at are (4), the expression for total transportation cost in terms of distance, weight and activity distribution, and (11), the areal constraint on activity distribution.

Note that (11) has the form of the capacity constraint in a measure-theoretic transportation problem, where the source space is (S, Σ_S, α) and the sink space is $(Q, \Sigma_q, ?)$, the question mark referring to an as-yet-unspecified requirement measure. Also (4) has the form of the objective function for this problem, ν being the unknown "flow" measure. The only missing ingredient is the requirement constraint, and this rôle will be filled by the "allotment" mentioned in footnote 5. ~~But we are now jumping ahead of ourselves.~~

B

Interpretations and Illustrations

Illustrations of theoretical concepts are always useful for making connections with the real world. For the "land-use" concept just developed they are especially important for two reasons. First, a great diversity of phenomena are encompassed by it, and this fact can be driven home only by examples. Second, the concept is unusual in several respects, and some of the associated terms — (such as "production" and "consumption") — are used in strange ways; all this needs elucidation.

~~First of all,~~ a land use is longitudinal, stretching over the entire time horizon. Suppose~~, for example,~~ that a site is successively vacant, used for farming, then for residing, manufacturing, office activities, and ends up as a parking lot. This whole succession (~~together~~ with the construction and demolition that occurs between phases) must be considered to be one land use, not a series of land uses. The production and consumption measures on $\underline{R} \times \underline{T}$, λ_1 and λ_2, will concentrate mass on different resource types in different epochs, of course, and the history of the "goings-on" could be reconstructed in part from a knowledge of these two measures.

We ~~Let us~~ examine these measures in more detail. In the entrepôt model all production is to be shipped to the nucleus. This means that we must include in "production" all resources that ~~which~~ leave the site and travel to the nucleus. Consider a residential land use in the context of an urban Thünen system, with the CBD as the nucleus. Any household member who makes a trip downtown, for work, shopping, recreation, or whatever, must be considered to be "produced" at that time and "exported" to the nucleus. The same is true for other household "exports" (e.g., outgoing mail and telephone calls, garbage and sewage~~, etc.~~) insofar as they move to a centralized processing point. Similarly, people traveling from the CBD to the household must be considered to be "consumed" at the time of the trip, and will be recorded in λ_2. The same is true for other resources coming in from downtown (consumer goods, water, gas, ~~and~~ electricity, incoming mail, ~~and~~ telephone calls~~, etc~~. A round trip counts both as an export and an import. Every trip must be counted separately.

What about <u>local</u> <u>trips</u> to neighborhood facilities (~~say~~ routine grocery shopping, children's school trips, local movies~~, etc~~? These should <u>not</u> be counted.[6] The basic principle for distinguishing these trips from those mentioned above is this: λ_1 and λ_2 are to be constructed so that the land-use weight, as determined by (2) and (3), is an accurate reflection of the "pull" of the nucleus on this land use. Extra trips to the CBD increase this pull; ~~that is~~ i.e., a land use with more such trips would save more in transport costs by moving one unit of (ideal) distance closer to the CBD than would a land use with fewer such trips, all other imports and exports being the same. But a change in local trips would be irrelevant in this respect.

925-10

A land use is defined by the triple $(\rho,\lambda_1,\lambda_2)$, and any variation in any of these measures, however slight, yields a different land use. Consider the general category of intensity variations, for example. ~~One can grow~~ Corn can be grown in a continuum of different ways, with variations of fertilizer input per acre leading to variations of corn yield per acre. Each of these different input-output level combinations is to be considered a different land use.

Intensity variations manifest themselves in the levels of inflowing and outflowing traffic per acre, and in the general degree of crowding of resources upon the site. One particular form that intensification takes is the phenomenon of multiple-story land uses, and this is important enough to deserve separate discussion.

An N-story structure provides a stack of N horizontal surfaces of support, on which N different processes can run simultaneously, one above the other.[7] This can be represented in ~~There are~~ at least two ways ~~of representing this~~ in terms of our categories. One approach identifies ~~Space,~~ S with supporting surfaces in general, including the (land?) surface of the earth and floors above (and possibly below) it. From this point of view, land uses are inherently "single-story". Most are placed at ground level, some on floors above or below ground. Multiple-story construction (including bridges, tunnels and pit mine construction) is then a way of creating new Space.

The second approach restricts ~~Space~~ space to the ~~Earth's~~ earth's surface. A site may be utilized for any of several land uses, some of which will be "multiple-story". The latter involve several processes stacked vertically, usually preceded by construction of the multiple-story structure ~~which~~ that supports them. For example, a ten-story office building, with detailed specification of what goes on at each floor, would be a typical multiple-story land use. On the preceding approach, it would decompose into ten separate land uses. ~~We shall~~, for the most part, we use the second approach.

Next, consider *time-displacement* as a form of variation among land uses; ~~For example~~ e.g., a trip is made sooner or later, a crop is harvested (and shipped) sooner or later. A special case is where the entire land use is shifted "rigidly" in time. To be precise, we must specify the structure of ~~Time,~~ T, as used in the model. Suppose ~~that~~ T is the nonnegative real numbers, so that the Thünen system is taken to begin at some moment, time zero, but unfolds indefinitely into the future. Land use q' is then said to be a *t_o-forward displacement of q* ($t_o > 0$) iff, for all $E \in \Sigma_r \times \Sigma_t$,

6½ pt

$$\lambda_i(q', E) = \lambda_i\left[q, \left\{(r,t) \,\middle|\, t \geq 0, \text{ and } (r, t + t_o) \in E\right\}\right], \quad (i = 1,2).$$

12 pt braces brackets vertical bar

That is, the production and consumption assigned to any measurable subset of $R \times T$ by q is the same as that assigned by q' to that set displaced forward t_o time units. Also, the "$t \geq 0$" insures that q' neither produces nor consumes before moment t_o.

This rigid displacement might arise in a land speculation situation, in which the controller of a site knows what he wants to do ~~with it,~~ but is waiting for the right moment to initiate operations. Just as with intensity variations, displacements are to be considered as different land uses.

What can be said about the weights of these various land uses? To find w(q) one needs the measures λ_1, λ_2 associated with q, as well as the ideal weight functions g_{in}, g_{out}, and then uses formulas (2) and (3). Certain general observations concerning procedures and "tendencies" are in order.

First ~~of all,~~ λ_1 and λ_2 are the production and consumption on one "ideal acre" of ~~Space.~~ Hence in measuring shipments to and from some actual land use, one must always adjust for this by dividing by the number of "ideal acres" on the site. As a first approximation one may identify ideal area with physical area, adjusting the former downward for sites with rough topography or poor drainage. In computing areas occupied by land uses, the accoutrements such as landscaped grounds and parking facilities should be included ~~counted in.~~ This ~~of course~~ will diminish the computed land-use weight by increasing the denominator.

G25-11 In general, the more "intensive" land uses tend to have greater weights, since λ_1 and λ_2 are larger. In particular, the weight of multiple-story land uses tends to rise with the number of stories. The imports and exports of such a land use are the sums of the imports and exports originating on the various floors. (~~For example~~ e.g., the trips to and from an office building are the sums of those terminating on the first, second, third, etc., floors). The ideal area of the site occupied by such a land use, ~~on the other hand~~ however, is that of the "ground floor" only, not the sum of the floor areas of the successive stories. Extremely high weights can thus be obtained via skyscraper construction.

As for time-displacements, ~~there will be~~ some tendency exists for forward displacement to make land uses lighter. This is a reflection of the tendency ~~already discussed~~ for ideal weights of resources to become lighter over time, owing to transportation improvements and discounting. Forward displacement shifts the masses distributed by λ_1 and λ_2 toward smaller values of the integrands g_{in} and g_{out}, reducing the integrals (2).

Turning attention to the ideal weight functions, we note the general tendency for the ideal weight/physical weight ratio to be higher for people than for non-human resources, since ~~This arises from the fact that~~ people demand more in the way of roominess, comfort, etc., for their own travel than ~~they demand~~ for the shipment of their chattels. Thus trips by people are an important contributor to the weight of most land uses, and probably dominate in those ~~land uses~~ involving "facilities", such as residences, churches, schools, hospitals, office buildings, and retail trade.

Ideal weight varies considerably from person to person. To assess what is involved ~~here~~, remember that "transportation cost" is a composite money valuation of many diverse components~~,~~ viz. not only fares and fuel consumption, but the value of time spent in traveling, risk of accident, discomfort and fatigue, etc. The potential traveler ~~himself~~ evaluates these dimensions in dollar terms, and it is this personal assessment ~~which~~ that constitutes his transport cost and determines his ideal weight.[8,9] Thus we may expect idiosyncratic elements to enter into ideal weight: Someone with a pathological fear of travel accidents will be very "heavy" on that account.

At the same time we may expect some regularities. Valuation of elapsed time will rise with foregone earnings, so that people with high wages (actual or imputed) will tend to be "heavy". Rich people will, on the average, be willing to pay more to avoid the same degree of accident risk and uncomfortable travel conditions than ~~will~~ poor people. Thus we may expect that ideal weight will rise with both earned and unearned income, and more so per dollar of earned than of unearned income.

~~There is,~~ however, one factor ~~which~~ goes counter to this tendency; viz. ~~namely,~~ ~~that~~ the rich tend to use speedier and more comfortable modes of transportation (e.g. airplanes, taxis, ~~and~~ private automobiles)~~, for example.~~ The automobile functions as a general "map shrinker" or, better, as a "levitator", reducing the ideal weights of those who customarily travel with it.

Ideal weight also varies over time for the same person. We have already discussed some general features of this time dependency. Over and above these are variations induced by changes in the opportunity value of ~~one's~~ time. Thus, weight probably rises when ~~one~~ a person enters the labor force, and falls at retirement. In the shorter run, weight is lower during evenings and weekends, when there are fewer earning opportunities.[10]

These factors all enter into the computation of the ideal weight functions, g_{in} and g_{out}, which in turn enter into the computation of land-use weight via (2) and (3). If one ~~is~~ deal~~ing~~s with a land use that is roughly steady-state over a long period ~~of time, and~~ in which trips by people are the dominant weight influence, the following schematic may be helpful for measurement purposes:

$$\text{weight of land use} = \frac{1}{i} \cdot \left[\begin{array}{c}\text{mean ideal}\\ \text{weight of}\\ \text{triptakers}\end{array}\right]\left[\begin{array}{c}\text{nuclear}\\ \text{trips per}\\ \text{person}\\ \text{per year}\end{array}\right]\left[\begin{array}{c}\text{population}\\ \text{density on}\\ \text{site}\end{array}\right]\left[\frac{\text{real acres}}{\text{ideal acres}}\right].$$

reduce if necessary to fit one line

9-25-12

Each of the right-hand factors should be roughly estimable. Here i is the discount rate, inserted to convert the flow to a present value; mean ideal weight is based on income, car ownership, etc., and is an average weighted by triptaking propensities; a round trip counts as two trips; the real/ideal areal ratio is based on topography, drainage, etc.

Finally, let us ~~take~~ note ~~of~~ the realism, or lack thereof, of the land-use concept we are using. The main departure from realism appears to lie in the absence of restrictions on the possible assignments ν (other than the areal capacity limitation (11)). Thus land uses can be mixed freely, and the presence of a distribution of uses in one region has no effect on what is feasible in an adjoining disjoint region. In short, neighborhood effects are excluded, as are the associated effects of "scale" and "indivisibility". As discussed in ~~chapter 4~~ section 4.6, the resulting departure from realism tends to be more severe, the smaller the scale of the system under discussion.

Another unrealistic simplification arises in the form of the areal constraint (11) itself. ~~As discussed above,~~ this says in effect that the demand-for-"room" function k(s,q) is factorable. In more picturesque language, no location has a <u>comparative</u> <u>advantage</u> over any other in relative suitability for any pair of land uses. It is easy to find exceptions; e.g. ~~For example~~, soil fertility is relevant for agricultural land uses but irrelevant for most urban land uses; hence <u>in</u>fertile land has a comparative advantage for the latter. Marshy land has a comparative advantage for certain kinds of recreational ~~land~~ uses, hilly land for residences, etc. On the institutional side, <u>zoning</u> is, in, effect the artificial introduction of comparative advantages by differential exclusions of certain land uses from certain regions. Some (but not all) forms of <u>real-estate taxation</u> have the same effect. All these phenomena are excluded by assumption. (Later we ~~shall~~ discuss the modifications induced by introducing some of them.)

A similar difficulty arises if the Thünen system starts up from some designated "time zero". If "time zero" precedes the settlement of the region one has only the geographic non-uniformities of nature to contend with. But if one places "time zero" <u>in</u> <u>medias</u> <u>res</u>, with a preceding period of settlement, ~~a~~ further complications ~~departure from the model~~ occur~~s~~. Man himself creates differential advantages, by building different structures in different places, ~~and~~ by leaving other places vacant, and by distributing himself non-uniformly over the landscape. This point is important, because the model has variables that refer to "time zero" (e.g. land values at that time) and not to other times.

A

8.5. The Allotment-Assignment Problem

We have ~~now~~ set up an apparatus of concepts for Thünen systems; ~~and~~ it is now ~~high~~ time to produce some models ~~for them~~. Two kinds ~~of models~~ will be considered. One ~~kind~~ is behavioral, the interactions of many agents in the real-estate market leading to the Thünen configuration of land uses. The other ~~kind~~ involves optimization, specifically, the minimization of total transport cost over a certain set of possible assignments. This again leads to the same land-use pattern, so that the free market interaction of numerous agents leads to the minimization of transport cost.

In the theory of urban structure, a long controversy has raged on exactly this point: Is the metropolis laid out ~~so as~~

to minimize the "friction of space" (and <u>should</u> it be so laid out)? The literature has been ably reviewed by Alonso, who concludes that "friction" is not (and should not be) minimized, because other desiderata — such as roominess — are also important.[11] While this is perfectly correct, the issue is not settled, because the meaning of "minimization" is left unclear. Specifically, one must name the <u>set of alternatives</u> under consideration before one can say that the alternative actually chosen does or does not minimize a certain objective. In the following development, the set of alternatives is such that the free market <u>does</u> minimize total transport cost over that set. (Whether it <u>should</u> do so is something we discuss later).

25-13 We first present the optimization model. The problem is to choose an assignment ν — which is, formally, a measure over the product space $(S \times Q, \Sigma_s \times \Sigma_q)$ — out of the feasible set of such assignments. The objective is to minimize total transport costs on shipments to and from the nucleus. According to our previous analysis, this is given by (4.4) ~~(4) of section 4~~:

151 3D 14

$$\int_{S\times Q} h(s)w(q)\nu(ds,dq), \tag*{(8.5.1) $\cancel{(1)}$}$$

where $h{:}S \to$ reals and $w{:}Q \to$ reals are the ideal distance and land-use weight functions, respectively. These are assumed to be measurable. Actually, all our basic results are still obtained with a much more general objective function than (1). We need the following concepts.

ga1 26-1

Definition: Let f be a real-valued function whose domain is ~~two~~ 2-space (i.e., the plane, reals^2). Function f has *positive cross-differences* iff, for all real numbers x_1, x_2, y_1, y_2 such that $x_1 < x_2$ and $y_1 < y_2$, we have

83½

$$f(x_1, y_1) + f(x_2, y_2) > f(x_1, y_2) + f(x_2, y_1). \qquad (8.5.2)\ \cancel{(2)}$$

Function f has *non-negative cross-differences* iff the same condition holds with "$\geq$" replacing "$>$" in (2).

These definitions easily extend to the case where the domain of f is a rectangle

$$X \times Y, \qquad (8.5.3)\ \cancel{(3)}$$

10 X 27 / 12 TR / #=2 ems

X and Y being real intervals. Simply restrict x_1, x_2 to lie in X and y_1, y_2 in Y. Now consider the integral

14 8½

$$\int_{S\times Q} f(h(s), w(q))\, v(ds, dq), \qquad (8.5.4)\ \cancel{(4)}$$

where f is a measurable function having positive (or perhaps non-negative) cross-differences. (From this point ~~on~~ we no longer write Q in boldface, since we are dealing with an abstract problem in which S and Q enter symmetrically.) Integral (1) is the special case of (4) in which f is simply the product: $f(x,y) = xy$. (This function clearly satisfies (2).) Hence any general results obtained using (4) as the objective function will apply to (1) in particular. The domain of f in (4) will usually be the plane, but, if the ranges of h and w are both bounded, it is possible — (and sometimes advantageous) — to let it be a rectangle with interval sides.[12]

Integrals (1) and (4) are written as definite integrals. In case they are infinite or not well-defined, however, we interpret them as indefinite integrals in the sense of pseudomeasures (ν being ~~sigma~~ σ-finite), and "minimization" of (1) or (4) is taken in the sense of (reverse) standard ordering of pseudomeasures. For well-defined finite integrals this of course reduces to the ordinary size comparison of definite integrals.

Next we come to feasibility conditions on ν. First, there is the areal capacity constraint (4.11) ~~(11) of section 4~~ (in equality form):

$$\nu(F \times Q) = \alpha(F), \qquad (8.5.5)\ \cancel{(5)}$$

for all $F \in \Sigma_s$. Here ideal area α is formally a measure on ~~Space,~~ (S, Σ_s). α is given and ~~is~~ assumed to be ~~sigma~~ σ-finite; (5) then guarantees that any feasible assignment ν will also be ~~sigma~~ σ-finite.

We now have two-thirds of a transportation problem, with objective function (4) and capacity constraint (5); ~~what is~~ is missing ~~is~~ the requirements constraint. It is ~~perfectly~~ possible to stop at this point and consider the "one-sided" transportation problem: Minimize (4) over measures ν, subject to (5). Formally, a model of this sort has been constructed by ~~Benjamin~~ Stevens (with an inequality constraint and in non-measure-theoretic terms).[13] ~~Note that~~ this "one-sided" problem is in fact the special case of the transportation problem (variant III) in which the requirement measure is zero. Hence the theory of the problem is more or less encompassed in the results of chapter 7.

In any case, this "one-sided" approach does not appear to get ~~one~~ very far, and for deep results one must go on to the full "two-sided" transportation problem. We ~~Let us~~ therefore add the following constraint:

$$\nu(\underline{S} \times \underline{G}) = \beta(\underline{G}), \qquad (8.5.6) \; \cancel{(6)}$$

for all $\underline{G} \in \Sigma_{\underline{q}}$. Here β is a given σ ~~sigma~~-finite measure on the space of land uses. β will be called the allotment measure and (6) the allotment constraint; the entire problem of minimizing (4) over assignments ν, subject to constraints (5) and (6), will be called the allotment-assignment problem.

Constraint (6) may be interpreted as follows. For any measurable set of activities $\underline{G}$, a total acreage allotment $\beta(\underline{G})$ is specified, which must be met by any assignment; ~~e.g.~~ two acres must be devoted to turnip growing, five acres to education, etc. There is still freedom to shuffle these land uses around over Space, but the totals are fixed. In contrast to the areal-capacity constraint (5), which represents a "real" restriction on possible assignments grounded in natural law or human institutions, (6) is best regarded as an "artificial" restriction added to attain certain results. (One exception: (6) is a "natural" restriction in layout problems, for which There is technology dictates the allotment, as in the separate processes of a manufacturing complex. But the Thünen framework is not well suited for layout problems.)[14]

Though artificial, in the sense of not representing an actual constraint on behavior, (6) ~~does~~ serves a function ~~which~~ that arises from the "inner logic" of Thünen systems. Consider the matter in the following light. Thünen systems arise in a ~~great~~ diversity of contexts, on all different scales. ~~What~~ they have in common ~~is~~ precisely the pattern of land uses: the ring structure and the ordering of uses. ~~What~~ they do not have in common ~~are~~ the particular land uses present in each, and their relative proportions: in short, the <u>allotments</u> of land uses. For someone looking for a universal theory of Thünen systems, the allotment measures are the contingent features. It is then reasonable to treat allotments as exogenous, and to set up a model ~~which~~ that yields the Thünen pattern of land uses regardless of what the allotment is. and (6) does just this: The allotment β is given <u>a priori</u>, and we are to find the optimal assignment <u>within</u> that given allotment. The resulting pattern is (within very wide limits) independent of β.

This approach is not used by any other contemporary model-builder in the Thünen tradition.[15] Rather, these authors try to predict the assignment of land uses without assuming the allotment in advance. In this sense our aim is narrower and more modest than theirs. But by the same token we cut through the aspects of these models which — (from the point of view of predicting the Thünen pattern) — are irrelevant and distracting, and thus we attain a deeper understanding of that pattern. From this point of view our assumptions are much weaker than any of theirs.

The allotment-assignment problem, then, is given by objective function (4) to be minimized subject to constraints (5) and (6) on assignments ν. This is, formally, a measure-theoretic transportation problem of variant *I* (~~that is~~ i.e., with equality constraints). (Other variants could be used, but *I* is the simplest.) The special feature of the allotment-assignment problem lies in the form of the integrand in (4), ~~especially~~ particularly in the fact that *f* has positive (or non-negative) cross-differences.

This special feature enables us to make very strong statements concerning the nature of the solution. We need a few concepts for this. First, on the plane it will be convenient to say that point (x_1, y_1) is <u>southwest</u> of (x_2, y_2) iff $x_1 < x_2$ and $y_1 < y_2$, <u>northwest</u> iff $x_1 < x_2$ and $y_1 > y_2$, etc. Next, given two subsets of the plane, E_1 and E_2, E_1 is said to be <u>southwest</u> of E_2 iff every point of E_1 is southwest of every point of E_2 in the sense just defined. Next, let functions $h{:}S \to$ reals and $w{:}Q \to$ reals be given, and let (s_1, q_1), (s_2, q_2) be two points of the cartesian product $S \times Q$; (s_1, q_1) is <u>southwest</u> of (s_2, q_2) iff $h(s_1) < h(s_2)$ and $w(q_1) < w(q_2)$. Finally, given two subsets of $S \times Q$, E_1 and E_2, E_1 is <u>southwest</u>

of E_2 iff every point of E_1 is southwest of every point of E_2 in this sense. Using this last concept we have the following:

Definition: Let ν be a measure on the product space $(S \times Q, \Sigma_s \times \Sigma_q)$, and let $h: S \to$ reals and $w: Q \to$ reals be functions. Measure ν satisfies the <u>measurable weight-falloff condition</u> iff there do not exist two sets $E_1, E_2 \in \Sigma_s \times \Sigma_q$, both of positive ν-measure, with E_1 southwest of E_2.

As with potentials, there is a corresponding topological concept. We suppose ~~that~~ topologies $\mathcal{T}_s$ and $\mathcal{T}_q$ have been placed on S and Q, respectively, making them topological spaces as well as measurable spaces. These determine a product topology, $\mathcal{T}_s \times \mathcal{T}_q$ on $S \times Q$, and this with $\Sigma_s \times \Sigma_q$ determines the support of measure ν.

Definition: Let ν be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$, and let $h: S \to$ reals and $w: Q \to$ reals be functions. Measure ν satisfies the <u>topological weight-falloff condition</u> iff there do not exist two points of support for ν, one of which is southwest of the other.

Roughly speaking, both these concepts state intuitively that h and w are negatively correlated: ν tends to concentrate mass where h is high and w low, and vice versa. Note that h and w enter symmetrically into these definitions, so that, instead of speaking of weight (w) falling off as distance (h) rises, one could speak of distance falling off as weight rises.

The two functions, h and w, determine a single function mapping $S \times Q$ into the plane, ~~namely~~ viz., the one ~~which~~ that assigns the value $(h(s), w(q))$ to the point (s,q). Assume that h and w are measurable; then this function is measurable. Hence, for any measure ν over universe set $S \times Q$, it induces a measure λ over the plane:

break only at =

$$\lambda(E) = \nu\{(s,q) \mid (h(s), w(q)) \in E\},$$

for any Borel subset E of the plane.

Now the two weight-falloff definitions above apply just as well to λ as to ν (the plane being furnished with its usual topology and Borel field, and with h and w each replaced by the identity map, $x \to x$, on the real line). Hence we have apparently four different concepts. But our next result shows that three of these conditions are logically equivalent.

#

D

Theorem: Let ν be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$, let $h{:}S \to$ reals and $w{:}Q \to$ reals be measurable, and let λ be the measure on the plane induced from ν by h and w. Then each of the following conditions implies the other two:

align right ()

(i) ν satisfies the measurable weight-falloff condition;

(ii) λ satisfies the measurable weight-falloff condition;

(iii) λ satisfies the topological weight-falloff condition.

#

D

Proof: (i) implies (ii): Let E_1, E_2 be two measurable subsets of the plane, with E_1 southwest of E_2. Their inverse images,

$$\{(s,q) \mid (h(s),\ w(q)) \in E_i\}$$ (8.5.7) ~~(7)~~

(i = 1, 2), retain this southwest-northeast relation. Hence at least one of them has ν-measure zero, which implies $\lambda(E_i) = 0$ for at least one E_i. Thus λ satisfies measurable weight-falloff.

(ii) implies (iii): Let z_1, z_2 be two points of the plane, with z_1 southwest of z_2. ~~There are~~ open discs E_1, E_2 about z_1, z_2 respectively exist such that E_1 is southwest of E_2. Sets E_1 and E_2 cannot both have positive λ-measure, hence z_1 and z_2 cannot both support λ. Thus λ satisfies topological weight-falloff.

(iii) implies (i): Let E_1, E_2 be two measurable subsets of $S \times Q$, with E_1 southwest of E_2, and consider their images in the plane:

$$F_i = \{(h(s),\ w(q)) \mid (s,q) \in E_i\},$$

i = 1, 2. F_1 is southwest of F_2. Hence at least one of these two sets – say F_j – cannot own any points of support for λ. Thus each point $(x,y) \in F_j$ has a measurable neighborhood of λ-measure zero. Now the usual topology of the plane has the strong Lindelöf property, so that F_j is contained in the union of a countable number of these neighborhoods. Call this union G; G is measurable, and $\lambda(G) = 0$. It follows that

$$0 = \nu\{(s,q) \mid (h(s),\ w(q)) \in G\} \geq \nu(E_j).$$ (8.5.8) ~~(8)~~

The equality in (8) arises from the fact that λ is induced from ν; the inequality arises from the fact that E_j is contained in the inverse image of F_j, which in turn is contained in the inverse image of G. Thus at least one of E_1, E_2 has ν-measure zero: ν satisfies measurable weight-falloff.

We now have a closed circle of implications. ||| □

We ~~shall~~ speak simply of ν (or λ) satisfying the weight-falloff condition in the event that any (hence all) of the above three conditions obtains.

What about the fourth condition, which is topological weight-falloff for ν? This depends on the topologies $\mathcal{T}_s$ and $\mathcal{T}_q$, which do not enter the definition of the other three concepts.

Theorem: Let ν be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$, let $h:S \to$ reals and $w:Q \to$ reals be measurable, and let λ be the measure on the plane induced from ν by h and w; also let $\mathcal{T}_s$ and $\mathcal{T}_q$ be topologies on S and Q, respectively; then

(i) if ν (or λ) satisfies the weight-falloff condition, and h and w are continuous functions, then ν satisfies the topological weight-falloff condition;

(ii) if ν satisfies the topological weight-falloff condition, and $\mathcal{T}_s \times \mathcal{T}_q$ has the strong Lindelöf property, then ν (or λ) satisfies the weight-falloff condition.

Proof: (i) If (s,q) is a point of support for ν, then $(h(s), w(q))$ ~~w(q))~~ is a point of support for λ. To show this, let $(h(s),w(q)) \in E_1 \subseteq E_2$, where E_1 is open and E_2 measurable. The inverse images (7) are open for E_1 and measurable for E_2, since h, w

are continuous and measurable. Hence the inverse image of E_2 has positive ν measure, implying $\lambda(E_2) > 0$. Thus $(h(s), w(q))$ supports λ.

Now let (s_i, q_i), $i = 1,2$, be two points of support for ν. Since $(h(s_i), w(q_i))$, $i = 1, 2$, are both points of support for λ, they cannot stand in a southwest-northeast relation. Hence neither can (s_i, q_i), $i = 1, 2$, so that ν satisfies topological weight-falloff.

¶ 18 (ii) Let E_1, E_2 be two measurable subsets of $S \times Q$, with E_1 southwest of E_2. At least one of these two sets, say E_j, cannot own any points of support for ν. Utilizing the strong Lindelöf property as in the preceding proof ((iii) implies (i)), it follows that E_j is contained in a ν-null set. Thus at least one of E_1, E_2 has ν-measure zero: ν satisfies weight-falloff. □

removes fl left here

These results imply that, if h and w are both continuous, and $T_s \times T_q$ has the strong Lindelöf property, then any of these weight-falloff conditions implies the other three. The next result establishes a connection between weight-falloff and allotment-assignment.

Theorem: Let (S,Σ_s,α) and (Q,Σ_q,β) be σ-finite measure spaces. Let $T_s \subseteq \Sigma_s$ and $T_q \subseteq \Sigma_q$ be topologies on S and Q, respectively. Let $h{:}S \to$ reals, $w{:}Q \to$ reals, and $f{:}\text{reals}^2 \to$ reals be functions such that the composite function $f(h(\cdot),w(\cdot)){:}S \times Q \to$ reals is measurable, and continuous with respect to $\check{T}_s \times \check{T}_q$. Finally, let f have <u>positive-cross-differences</u>.

Then, if measure ν° is <u>unsurpassed</u> for the allotment-assignment problem of minimizing (4) (reverse standard order) subject to the constraints (5) and (6), it follows that ν° satisfies the <u>topological weight-falloff condition</u> (with respect to h and w).

<u>Proof</u>: The premises imply that ν° satisfies the <u>circulation condition</u> (7.5.10) or (7.5.11) ~~((10) or (11) of 7.5)~~. Thus, if (s_1, q_1) and (s_2, q_2) are two points of support for ν°, we have

$$f_{12} + f_{21} - f_{11} - f_{22} \geq 0, \qquad \text{(8.5.9)} \; \cancel{(9)}$$

where f_{ij} abbreviates $f(h(s_i), w(q_j))$, $i, j = 1, 2$. We cannot have both $h(s_1) < h(s_2)$ and $w(q_1) < w(q_2)$, because in this case (9) would contradict the positive cross-differences condition (2). That is, (s_1, q_1) cannot be southwest of (s_2, q_2). Thus ν° satisfies topological weight-falloff. ▯

Our next result is similar to this one. Though its proof is more complicated, it is also more interesting because it makes no continuity assumptions; indeed, it uses no topological concepts whatever. Recall that a function is <u>half-bounded</u> iff it is bounded below or bounded above (or both, i.e., bounded).

<u>Theorem</u>: Let (S, Σ_s, α) and (Q, Σ_q, β) be σ-finite measure spaces. Let $h: S \to$ reals, $w: Q \to$ reals, f: reals$^2 \to$ reals be

functions such that f has positive cross-differences, and the composite function $f(h(\cdot), w(\cdot)): S \times Q \to$ reals ~~is~~ is measurable. Let measure ν° be best for the allotment-assignment problem of minimizing (4) (reverse standard order) subject to constraints (5) and (6).

Then ν° satisfies the (measurable) weight-falloff condition (with respect to h, w).

If ν° is merely unsurpassed, the same conclusion follows provided ~~one~~ we add the premise that ~~$f(h(\cdot), w(\cdot))$~~ f is half-bounded on any bounded subset of the plane, and that h, w are measurable.

Proof: Assume ~~that~~ ν° violates measurable weight-falloff, so that ~~there are~~ sets $F_1, F_2 \in \Sigma_s \times \Sigma_q$ of positive ν°-measure exist, with F_1 southwest of F_2. Either of these may in fact have infinite measure, but in any case they will contain subsets $G_i \subseteq F_i$, $i = 1, 2$, of positive finite measure, since ν° is ~~sigma~~ σ-finite. Define the measures ν_1, ν_2 on $(S \times Q, \Sigma_s \times \Sigma_q)$ by

$$\nu_i(H) = \nu^\circ(H \cap G_i)/\nu^\circ(G_i), \qquad (8.5.10)\ \cancel{(10)}$$

$i = 1, 2$, $H \in \Sigma_s \times \Sigma_q$, and define the signed measure ν by

$$\nu = (\nu_1' \times \nu_2'') + (\nu_2' \times \nu_1'') - \nu_1 - \nu_2. \qquad (8.5.11)\ \cancel{(11)}$$

(Here ν_i', ν_i'' are the left and right marginals, respectively, of ν_i). ν is well-defined, since both ν_1 and ν_2 are bounded measures. Finally, consider the (signed) measure

$$\nu^{o} + z\nu, \qquad (8.5.12)\ ~~(12)~~$$

where $z = \frac{1}{2}\min[\nu^{o}(G_1), \nu^{o}(G_2)] > 0$. One easily verifies that (12) remains feasible for the allotment-assignment problem; (cf. (7.5.20) ~~(20) of 7.5~~).

~~Now,~~ taking the case where ν^{o} is best, we ~~shall~~ reach a contradiction. Since (12) is feasible, we must have

$$\int f(h(s), w(q))\,\nu(ds,dq) \geq 0. \qquad (8.5.13)\ ~~(13)~~$$

(The integral in (13) is a pseudomeasure over $S \times Q$, and "$\geq$" refers to standard ordering; cf. (7.5.21) ~~(21) of 7.5~~). The integral in (13) can be written as the sum of four indefinite integrals, corresponding to the splitting of ν into its four components (11). We now show that

$$\int_{S\times Q} f\, d(\nu_1' \times \nu_2'') + \int_{S\times Q} f\, d(\nu_2' \times \nu_1'') - \int_{S\times Q} f\, d\nu_1 - \int_{S\times Q} f\, d\nu_2 \qquad (8.5.14)\ ~~(14)~~$$

is a well-defined expression. (Here "f" abbreviates $f(h(\cdot), w(\cdot))$.) That is, each of the four definite integrals in (14) is well-defined, and their sum is not of the form $\infty - \infty$. To see this, note that there are numbers x, y such that

$$h(s_1) \leq x \leq h(s_2) \qquad (8.5.15)\ (15)$$

~~and~~

$$w(q_1) \leq y \leq w(q_2), \qquad (8.5.16)\ ~~(16)~~$$

for all $(s_1, q_1) \in G_1$, $(s_2, q_2) \in G_2$, where one of the inequality signs in (15), and one in (16), can be replaced by

"<". This follows from G_1 being southwest of G_2. Condition (15) determines a partition of S into two pieces, one set satisfying the left, the other the right, inequality. Similarly, (16) splits Q into two pieces. Together these split $S \times Q$ into four pieces. Set G_1 is contained in the "southwest quadrant" of low h, w values; G_2 is contained in the "northeast quadrant" of high h, w values. It follows that the complement of the (southwest, northeast) quadrant has ν_1-, ν_2-measure zero, respectively. And from this it follows that the complement of the (southeast, northwest) quadrant has $\nu_2' \times \nu_1''$-, $\nu_1' \times \nu_2''$-measure zero, respectively. Thus the four components of ν are mutually singular in pairs. The indefinite integral (13) can therefore be expressed as a direct sum of four integrals over these quadrants. Being comparable to 0, the integral (13) must be a signed measure, so that (14) is indeed well-defined. In fact, the relation (13) implies that the expression (14) is nonnegative, by the standard integral theorem.

Now consider the "four-dimensional" product-measure space

$$(S_1 \times Q_1, \Sigma_s \times \Sigma_q, \nu_1) \times (S_2 \times Q_2, \Sigma_s \times \Sigma_q, \nu_2).$$

Here S_1 and S_2 are replicas of S, and Q_1 and Q_2 are replicas of Q; the subscripts are added for clarity. We have

$$\int_{S\times Q} f_i \, d\nu_i = \int_{S_1\times Q_1\times S_2\times Q_2} f_{ii} \, d(\nu_1 \times \nu_2), \qquad (8.5.17)\ \cancel{(17)}$$

$i = 1, 2$, and

23 31 105 78 86

$$\int_{S\times Q} f\, d(\nu_i' \times \nu_j'') = \int_{S_1\times Q_1\times S_2\times Q_2} f_{ij}\, d(\nu_1 \times \nu_2), \qquad (8.5.18)\ \cancel{(18)}$$

subscript to ∫

where $(i,j) = (1,2)$, and $(i,j) = (2,1)$.

In (17) and (18), "f" on the left again abbreviates $f(h(\cdot), w(\cdot))$, while f_{ij} stands for $f(h(s_i), w(q_j))$, i and j ranging over 1, 2 (four cases). The four equations in (17) and (18) all arise from the induced integrals theorem, resulting from four different projections from the space $S_1 \times Q_1 \times S_2 \times Q_2$ to $S \times Q$. Thus (17) for $i = 1$ arises from the projection $(s_1, q_1, s_2, q_2) \to (s_1, q_1)$; for $i = 2$ it arises from $(s_1, q_1, s_2, q_2) \to (s_2, q_2)$. Equation (18) for $(i,j) = (1,2)$ arises from $(s_1, q_1, s_2, q_2) \to (s_1, q_2)$; for $(i,j) = (2,1)$ it arises from $(s_1, q_1, s_2, q_2) \to (s_2, q_1)$. The only difficulty in demonstrating all this arises in (18), where it must be shown, e.g., ~~for example~~, that $\nu_1' \times \nu_2''$ is the measure induced from $\nu_1 \times \nu_2$ by the projection $(s_1, q_1, s_2, q_2) \to (s_1, q_2)$. This follows from direct substitution in the definition of product-measure. The well-definedness of the left integrals in (17) and (18) implies the well-definedness of the right integrals and the stated equalities.

1/28/6

Equations (17), (18), and the non-negativity of (14), then yield

55 78 240

$$\int_{S_1\times Q_1\times S_2\times Q_2} (f_{12} + f_{21} - f_{11} - f_{22})\, d(\nu_1 \times \nu_2) \geq 0. \qquad (8.5.19)\ \cancel{(19)}$$

subscript to ∫

fl left

But this is a contradiction. First ~~of all~~, $\nu_i((S_i \times Q_i)\backslash G_i)=0$, $i = 1, 2$, so that the complement of $G_1 \times G_2$ has $\nu_1 \times \nu_2$-measure zero. Second, $(\nu_1 \times \nu_2)(G_1 \times G_2) = \nu_1(G_1)\cdot\nu_2(G_2) = 1 > 0$. Finally, the integrand in (19) is negative on $G_1 \times G_2$, since f has positive cross-differences. Hence the integral in (19) must be negative, a contradiction. This proves the first half of the theorem.

Now take the case where ~~$f(h(\cdot), w(\cdot))$~~ f is half-bounded on bounded sets and ν° is merely unsurpassed, and again assume that ν° violates measurable weight-falloff. Proceeding as above, we find a set G_1 southwest of a set G_2, both with positive finite ν°-measure. Each of these contains a subset of positive measure on which $f(h(\cdot), w(\cdot))$, as well as h and w themselves, are ~~is~~ bounded, for the measurable sets

$$G_i \cap \{(s,q) \mid m \le f(h(s), w(q)) < m + 1, \ n \le h(s) < n+1, \ p \le w(q) < p+1\},$$

m, n, p ~~m~~ $= 0, \pm 1, \pm 2, \ldots$, countably partition G_i, hence one of these has positive measure. For simplicity, we designate these subsets by the same symbols, G_1 and G_2.

~~Now~~ define ν_1 and ν_2 as in (10) and consider the expression (14). The four measures appearing in these integrals are bounded. The complement of G_i has ν_i-measure zero, $i = 1, 2$, and $f(h(\cdot), w(\cdot))$ is bounded on $G_1 \cup G_2$; hence the last two integrals in (14) are well-defined and finite. As for the first two, ~~the fact that $f(h(\cdot), w(\cdot))$ is half-bounded implies~~ $\nu_i' \times \nu_j''$ are both zero off some set on which h and w are bounded, and the integrands are half-bounded on this set. It follows that they, too, are well-defined, and not infinite of opposite sign. Hence the whole expression (14) is well-defined.

It follows that the indefinite integral (13) is a signed measure, not a proper pseudomeasure, and is therefore comparable to 0 under standard ordering. ~~Now~~ Just as above we construct the new feasible solution (12), and the unsurpassedness of ν° together with comparability then yields relation (13). Hence (14) is nonnegative. The argument above then again yields a contradiction, and the last part of the theorem is proved. ||| $\square$

One immediate generalization: The premises on f need hold only for f restricted to the range of (h,w). The proof above still holds verbatim.

We make

These theorems have a very simple intuitive meaning. Suppose ν fails to satisfy, say, topological weight-falloff, so that there are points of support $(\underline{s}_1, \underline{q}_1)$, $(\underline{s}_2, \underline{q}_2)$, the first southwest of the second. Then shift a mass of activities in the "neighborhood" of $\underline{q}_1$ from the "neighborhood" of location $\underline{s}_1$ to the "neighborhood" of location $\underline{s}_2$, and shift a mass of activities in the "neighborhood" of $\underline{q}_2$ in the opposite direction. This reshuffling does not affect the feasibility conditions (5) and (6), and the positive cross-difference condition on $\underline{f}$ implies that the total transport cost (4) has been reduced. Hence the original assignment ν has been surpassed. The proofs above are merely a rigorization of this informal argument.

If we are looking for optimal solutions to the allotment-assignment problem under the mild conditions stated above, these theorems say ~~s that~~ we might as well confine our attention to assignments ν satisfying some weight-falloff condition. But

~~note that~~ there is as yet no guarantee that such solutions will be best, or even unsurpassed. We do not know at this stage whether ~~there~~ even exist such feasible assignments. And even if they ~~exist~~ do, they may not be optimal, since the possibility remains that there are no optimal solutions.

We ~~shall~~ attack these difficulties by transforming the original allotment-assignment problem into a simpler one. Specifically, we "induce" the original problem, which is set in the product space $S \times Q$, into the plane by means of the functions h and w. measure α on (S, Σ_s) is induced by h into a measure on the real line. Similarly, β on (Q, Σ_q) is induced by w into a measure on the real line. Finally, ~~we have already mentioned that~~ ν induces a measure λ on the plane via the combined function $(s,q) \to (h(s), w(q))$.

We ~~shall~~ retain the notation α, β for the measures on the real line induced by these respective original measures, and rely on context to distinguish them. The transformed allotment-assignment problem now reads as follows:

Find a measure λ on the plane ~~which~~ that satisfies the constraints

$$\lambda'(E) = \alpha(E) \qquad (8.5.20)\ \cancel{(20)}$$

~~and~~

$$\lambda''(E) = \beta(E) \qquad (8.5.21)\ \cancel{(21)}$$

for all Borel sets E on the real line, and that minimizes

$$\int_{\boxed{\text{reals}^2}} f\, d\lambda. \qquad (8.5.22)\ \cancel{(22)}$$

188 ? 34 36 subscript to 5

fl lyt

Here λ', λ'' are the left- and right-marginals of λ respectively, and are ~~of course~~ measures on the real line. Constraints (20) and (21) are the analogues of the areal-capacity and allotment constraints, (5) and (6), respectively. The integrand f in the objective function (22) is the same as the f appearing in (4), and so has positive (or non-negative) cross-differences.

The transformed measures α and β in (20) and (21) must be σ ~~sigma~~-finite. This is not implied by the σ ~~sigma~~-finiteness of the original α, β measures in (5) and (6), and must be explicitly postulated. In fact we ~~shall~~ make an even stronger assumption below.

This transformed problem has been placed on the plane. More generally, it could be placed on a rectangle $X \times Y$ (with interval sides) provided only that the ranges of h and w are contained in X, Y, respectively. $X \times Y$ is then the domain of f and the universe set of λ; X and Y are the universe sets of transformed α and β, respectively.

The objective function (22) is written as a definite integral. But, just as with (4), if it is not well-defined or finite for certain feasible measures λ, it is to be interpreted as an indefinite integral pseudomeasure, and "minimization" is to be understood in the sense of (reverse) standard ordering.

The first thing to notice about this transformed problem is that, formally, it is just a special case of the allotment-assignment problem: Both (S,Σ_s) and (Q,Σ_q) are the real line with its Borel field, and both h and w are the identity function.

26-8 The preceding theorems then apply and take a very simple form:

> Let measurable f:reals$^2 \to$ reals have ~~a~~ positive cross-differences, and let λ° be best for the problem of minimizing (22) subject to (20) and (21); then λ° satisfies the weight-falloff condition. The same conclusion holds if λ° is merely unsurpassed, provided f is ~~either continuous or~~ half-bounded on bounded sets.

18 We ~~shall~~ now investigate the feasibility and optimality relations between the original and the transformed allotment-assignment problems. The following property of induced pseudomeasures is needed.

Lemma: Let (B,Σ') and (C,Σ'') be measurable spaces, and $g:B \to C$ a measurable function. Let ν, ν° be two measures on (B,Σ'), and λ, λ° the measures on (C,Σ'') induced by g from ν, ν°, respectively, all four of these measures being σ-finite.

Then (λ°,λ) is the pseudomeasure induced by g from pseudomeasure (ν°,ν).

Proof: Recall (~~page~~ 000) that the pseudomeasure induced by g from ψ on (B,Σ') is (μ_1,μ_2), where μ_1, μ_2 are the measures induced by ψ^+, ψ^-, respectively; this is defined iff μ_1, μ_2 are both σ-finite. ~~Now~~ For $\psi = (\nu^\circ,\nu)$ we have

$$\psi^+ + \nu = \psi^- + \nu^\circ \qquad (8.5.23)$$

fl left

(equivalence theorem). This implies

$$\mu_1 + \lambda = \mu_2 + \lambda^{\circ}. \qquad (8.5.24) \quad \cancel{(24)}$$

To show this, let $E \in \Sigma''$, and apply (23) to $\{b \mid g(b) \in E\}$. The four terms in (23) are respectively equal to the four terms in (24) applied to E. Hence (24) is true.

Now $\psi^+ \leq \nu^{\circ}$ and $\psi^- \leq \nu$ (minimizing property of the Jordan form). Hence $\mu_1 \leq \lambda^{\circ}$ and $\mu_2 \leq \lambda$, so that μ_1 and μ_2 are ~~sigma~~ σ-finite, and the induced pseudomeasure exists. From (24) we obtain

break only at =

$$(\mu_1, \mu_2) = (\lambda^{\circ}, \lambda),$$

by the equivalence theorem again, ~~which completes the proof.~~ |||

D

Theorem: Let (S, Σ_s, α) and (Q, Σ_q, β) be measure sp~~a~~aces, and let $h: S \to$ reals, $w: Q \to$ reals, and $f:\text{reals}^2 \to$ reals be measurable functions. Let α and β, as well as their namesakes induced on the real line by h and w, respectively, be ~~sigma~~ σ-finite. Let ν° be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$ ~~which~~ that is feasible for the original allotment-assignment problem, (4), $\frac{1}{N}$ ~~(5),~~ (6), and let λ° be the measure on the plane induced from ν° by the mapping $(s,q) \to (h(s), w(q))$. Then

align right ( )

(i) λ° is feasible for the transformed problem, (20) $\frac{1}{N}$(22) ~~, (21),~~ ~~(22),~~ ; ~~and~~

(ii) if λ° is unsurpassed for the transformed problem, ~~then~~ ν° is unsurpassed for the original problem.

Proof: (i) The feasibility condition on ν° is that its left- and right-marginals coincide with α and β respectively. For any Borel set E on the real line we have

$$\lambda^{\circ\prime}(E) = \lambda^\circ(E \times \text{reals}) = \nu^\circ\left(\{s \mid h(s) \in E\} \times Q\right) = \alpha\{s \mid h(s) \in E\}. \qquad \text{(8.5.25)} \; \cancel{(25)}$$

The right-hand term in (25), however, is simply $\alpha(E)$ for the induced measure α; this proves (20) for λ°. A similar argument establishes (21). Thus λ° is feasible for the transformed problem.

(ii) Assuming that ν° is surpassed, we ~~shall~~ prove that λ° is surpassed. Abbreviate the composite function $f(h(\cdot), w(\cdot))$: $S \times Q \to$ reals by k. Then by hypothesis ~~there~~ exists a feasible ν such that

$$\int (-k)\, d\nu > \int (-k)\, d\nu^\circ. \qquad \text{(8.5.26)} \; \cancel{(26)}$$

(These are indefinite integrals over $S \times Q$, and ">" is the "greater than" relation for standard ordering of pseudomeasures. The minus sign is introduced to convert the objective from minimization to maximization.)

Relation (26) is equivalent to

$$\int k\, d\psi > 0, \qquad \text{(8.5.27)} \; \cancel{(27)}$$

where ψ is the pseudomeasure (ν°, ν). From the definition of standard order, (27) is the same as

29] $$\int_{S\times Q} k^{+}d\psi^{+} + \int_{S\times Q} k^{-}d\psi^{-} > \int_{S\times Q} k^{+}d\psi^{-} + \int_{S\times Q} k^{-}d\psi^{+}, \qquad \text{12 (8.5.28)} \; \cancel{(28)}$$

where these are four ordinary definite integrals.

~~Now~~ let μ_1, μ_2 be the measures induced on the plane from ψ^{+}, ψ^{-}, respectively, by the mapping $(s,q) \to (h(s), w(q))$. Then (28) implies

30] $$\int_{\text{reals}^2} f^{+}d\mu_1 + \int_{\text{reals}^2} f^{-}d\mu_2 > \int_{\text{reals}^2} f^{+}d\mu_2 + \int_{\text{reals}^2} f^{-}d\mu_1, \qquad (8.5.29) \; \cancel{(29)}$$

since by the ordinary induced integrals theorem, the four integrals in (29) are equal to the integrals in (28), respectively from left to right.

CONT. 26-9 500 togo

We have $\psi^{+} \leq \nu^{o}$ by the minimizing property of the Jordan form. Hence their induced measures stand in the same relation: $\mu_1 \leq \lambda^{o}$. A similar argument yields $\mu_2 \leq \lambda$, where λ is the measure on the plane induced from ν. Part (i) established that λ^{o} and λ were feasible for the transformed problem. Hence they, and therefore μ_1 and μ_2, are σ-finite. (29) then implies

$$\int f \, d(\mu_1,\mu_2) > 0, \qquad (8.5.30) \; \cancel{(30)}$$

in terms of pseudomeasure (μ_1,μ_2). This latter is the pseudomeasure induced from $\psi = (\nu^{o},\nu)$, and we now invoke the preceding lemma to establish the pseudomeasure equality:

$$(\mu_1, \mu_2) = (\lambda^{o},\lambda). \qquad (8.5.31) \; \cancel{(31)}$$

Finally, (30) and (31) yield

$$\int_{\wedge}(-f)_{\wedge}\, d\lambda \;>\; \int_{\wedge}(-f)_{\wedge}\, d\lambda^{o}.$$

But $\int_{\wedge}(-f)_{\wedge}\, d\lambda$ is just the (negated) objective function (22) for the transformed problem, ~~and~~ so λ^{o} is surpassed by λ.

(~~Note that~~ if (4) is a well-defined, finite, definite integral for all feasible ν, ~~then~~ part (ii) of this theorem can be proved in a few lines, since ~~For then~~ (22), for the λ induced from ν, is equal to (4) by the induced integrals theorem. ~~Also,~~ in this simple case, the distinction between "unsurpassed" and "best" disappears.)

In general, these results cannot be strengthened. One cannot infer the feasibility of ν^{o} from that of λ^{o}, nor the optimality (in any sense) of λ^{o} from that of ν^{o}. This latter inference, for example, is blocked by the following difficulty. To establish the optimality of λ^{o} one must consider all other feasible measures λ. But it is not necessarily the case that every such λ is the induced measure from some feasible ν. (In fact there may not be any such ν, feasible or not.) The optimality of ν^{o} tells nothing about such "uninduced" feasible measures λ, so that the optimality of λ^{o} cannot be inferred.

We ~~shall~~ now make a fairly detailed study of the transformed problem, and then use the preceding theorems to draw conclusions about the original problem.

Consider the following conditions on the original measures α and β:

P

Definition: Given measure α on (S, Σ_s), and measurable $h: S \to$ reals, α is *finite from below* iff

$$\alpha\{s \mid h(s) < x\} \text{ is finite,} \qquad (8.5.32)\ \cancel{(32)}$$

for all real numbers x. Similarly, given β on (Q, Σ_q) and measurable $w: Q \to$ reals, β is *finite from above* iff

$$\beta\{q \mid w(q) > x\} \text{ is finite,} \qquad (8.5.33)\ (33)$$

for all real x.

The interpretation of (32) is that the ideal area of the region *within* ideal distance x of the nucleus is finite, for any real x. This is not implausible, and does not preclude the possibility that the ideal area of Space as a whole is infinite. Similarly, (33) states that the allotment to the set of land uses of weight exceeding x is finite, for any real x.

These properties can be stated in logically equivalent form in terms of the transformed measures α and β. Namely, (32) and (33) are the same as

$$\alpha\{y \mid y < x\} \text{ and } \beta\{y \mid y > x\} \text{ are finite,} \qquad (8.5.34)\ \cancel{(34)}$$

for all real x, respectively. That is, the α-measure of any left half-line, and the β-measure of any right half-line, are finite. Note that (34) implies the ~~sigma~~ σ-finiteness of α and β.

~~The reason for imposing~~ these conditions are imposed because ~~is that~~ they insure the existence of a unique measure λ on the plane satisfying the weight-falloff condition with marginals α and β. As a preliminary, we ~~let us~~ establish a connection between the weight-falloff and northwest corner conditions. We have already encountered the latter condition for a measure λ on a product space $A \times B$, where A and B are both countable (~~in~~ Section 7.2). The general idea is that, given its marginals, λ have as much mass as possible concentrated into "corner" sets, these being defined in terms of certain complete orderings on A and B. In the present case, both A and B are the real line, which has a natural order. λ is then to concentrate its mass in the "corner" with low A-values and high B-values. The following definition makes this precise.

Definition: Let λ be a measure on the plane, with left and right marginals λ', λ'', respectively; λ satisfies the northwest corner condition iff

12pt brackets

$$\lambda\{(x,y) \mid x < x_1,\ y > y_1\} = \min\left(\lambda'\{x \mid x < x_1\},\ \lambda''\{y \mid y > y_1\}\right) \qquad (8.5.35)\ \ \cancel{(35)}$$

for all pairs of real numbers (x_1, y_1).

The set $\{(x,y) \mid x < x_1,\ y > y_1\}$ is the quadrant of the plane "northwest" of the point (x_1, y_1). It can be seen ~~is easy to see~~ that, for any measure, the left side of (35) never exceeds the right. Hence (35) is indeed the condition that the mass on these northwest sets be as large as possible.

Theorem: Let λ be a measure on the plane whose left-marginal $\lambda' = \alpha$ is finite from below, and whose right marginal $\lambda'' = \beta$ is finite from above; i.e., ~~(that is,~~ (34) holds. Then λ satisfies the weight-falloff condition iff it satisfies the northwest corner condition.

Proof: For any point (x_1, y_1) define the three sets

$$E = \{(x,y) \mid x < x_1,\ y > y_1\},$$

$$F = \{(x,y) \mid x < x_1,\ y \le y_1\}, \qquad (8.5.36)$$

$$G = \{(x,y) \mid x \ge x_1,\ y > y_1\}.$$

Then

$$\lambda(E) + \lambda(F) = \lambda'\{x \mid x < x_1\}, \qquad (8.5.37)$$

~~and~~

$$\lambda(E) + \lambda(G) = \lambda''\{y \mid y > y_1\}, \qquad (8.5.38)$$

so that (35) takes the form

12 pt bracket

$$\lambda(E) = \min\big(\lambda(E) + \lambda(F),\ \lambda(E) + \lambda(G)\big). \qquad (8.5.39)$$

Now let λ satisfy weight-falloff. Then either $\lambda(F) = 0$ or $\lambda(G) = 0$, since F is southwest of G. Either of these cases yields (39), so that λ satisfies northwest corner.

Conversely, let (39) be true for all (x_1, y_1). By (34), all terms in (37) and (38) are finite. (39) then implies that either $\lambda(F) = 0$ or $\lambda(G) = 0$. Now let (x_2, y_2) and (x_3, y_3) be any two points, the first southwest of the second, and choose

x_1 and y_1 such that $x_2 < x_1 < x_3$, $y_2 < y_1 < y_3$. Then F and G of (36) are measurable neighborhoods of (x_2,y_2) and (x_3,y_3), respectively, so these points cannot both support λ. This proves that λ satisfies weight-falloff. ||||

Our next result is based on the theory of distribution functions. Owing to the fact that we are dealing with "northwest" rather than "southwest" sets, the standard theorem (~~page~~ 000) must be rephrased in a slightly different form:

Lemma: Let g:reals$^2 \to$ reals satisfy the following three conditions:

(i) for all real numbers x_1, x_2, y_1, y_2 with $x_1 < x_2$ and $y_1 < y_2$ we have

$$g(x_1,y_1) + g(x_2,y_2) \le g(x_1,y_2) + g(x_2,y_1); \qquad (8.5.40)\ \cancel{(40)}$$

(ii) g is continuous from the "northwest": for any (x_1,y_1) and any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$|g(x,y) - g(x_1,y_1)| < \varepsilon \qquad (8.5.41)\ \cancel{(41)}$$

for any (x,y) satisfying: $x_1 - \delta \le x \le x_1$ and $y_1 + \delta \ge y \ge y_1$;

(iii) for fixed y, $g(x,y) \to 0$ as $x \to -\infty$, and, for fixed x, $g(x,y) \to 0$ as $y \to +\infty$.

Then there is exactly one measure λ on the plane satisfying

$$\lambda\{(x,y) \mid x < x_1,\ y > y_1\} = g(x_1,y_1) \qquad (8.5.42)\ \cancel{(42)}$$

for all points (x_1,y_1).

9a) 527-1

fl left

10/12 X 27 TR #=2ems

Note the opposite sign orientations of x and y in parts (ii) and (iii). Condition (40) states that g has non-positive cross-differences (cf. (2)), whereas the usual distribution functions have the opposite property. This lemma follows from the usual statement by "reflecting" the plane through the X-axis: $(x,y) \to (x,-y)$. We are now ready for the main result.

Theorem: Let α and β be two measures on the real line L satisfying (34); α is finite from below and β finite from above. Also let $\alpha(L) = \beta(L)$. Then there is exactly one measure λ on the plane having α and β as its left and right marginals, and satisfying the weight-falloff condition.

Proof: Define g:reals^2 $\to$ reals by

$$g(x_1,y_1) = \min\left(\alpha\{x|x < x_1\},\ \beta\{y|y > y_1\}\right) \qquad (8.5.43)$$

This is indeed finite, ~~real-valued~~, by (34). We now show that g satisfies (i), (ii), and (iii) of the preceding lemma.

18 ~~As for~~ (i), choose real numbers $x_1 < x_2$ and $y_1 < y_2$. We clearly have $g(x_1,y_1) \le g(x_2,y_1)$ and $g(x_2,y_2) \le g(x_2,y_1)$. If $\alpha\{x|x < x_1\} \le \beta\{y|y > y_2\}$, then also $g(x_1,y_1) = g(x_1,y_2)$. If $\alpha\{x|x < x_1\} \ge \beta\{y|y > y_2\}$, then also $g(x_2,y_2) = g(x_1,y_2)$. In either case, the equality, combined with one of the inequalities, yields (40).

numbers here fl l

18 ~~As for~~ (ii), note that $\alpha\{x|x < x_1\}$, as a function of x_1, is continuous from below; and $\beta\{y|y > y_1\}$, as a function of y_1,

is continuous from above. (This follows from the basic continuity property of measures, ~~page~~ 00§).) That is, given (x_1, y_1), for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|\alpha\{x|x < x_2\} - \alpha\{x|x < x_1\}| < \varepsilon \qquad (8.5.44)\ \cancel{(44)}$$

~~and~~

$$|\beta\{y|y > y_2\} - \beta\{y|y > y_1\}| < \varepsilon, \qquad (8.5.45)\ \cancel{(45)}$$

for all x_2, y_2 satisfying: $x_1 - \delta \leq x_2 \leq x_1$ and $y_1 + \delta \geq y_2 \geq y_1$. Then (44) and (45) together yield (41).

¶ ~~As for~~ (iii), the limit of $\alpha\{x|x < x_1\}$ is zero as $x_1 \to -\infty$, and the limit of $\beta\{y|y > y_1\}$ is zero as $y_1 \to +\infty$. Hence the limit of $g(x_1, y_1)$ is zero in both cases, which is (iii).

¶ Applying the lemma, we conclude that ~~there exists~~ a measure λ exists satisfying (42). We now show that this λ has the required properties.

Let y_1 go to $-\infty$ in (42). The left side has the value $\lambda'\{x|x < x_1\}$ as limit, where λ' is the left marginal of λ. $\beta\{y|y > y_1\}$ has the value $\beta(L) = \alpha(L)$ as limit, which is at least as large as $\alpha\{x|x < x_1\}$ for any x_1. Hence $g(x_1, y_1)$ approaches $\alpha\{x|x < x_1\}$ as limit. This proves that

$$\lambda'\{x|x < x_1\} = \alpha\{x|x < x_1\}$$

for all real x_1; λ' and α have the same distribution function, so and must therefore coincide.

Letting x_1 go to $+\infty$ in (42), a similar argument shows that $\lambda'' = \beta$. Thus α and β are indeed the left and right marginals of λ, respectively.

This being the case, (42) is the same as (35), so that λ satisfies the northwest corner condition. By the preceding theorem, it therefore satisfies weight-falloff.

The existence of λ has now been established. To prove uniqueness, let λ satisfy weight-falloff and have marginals α and β. By the preceding theorem, λ satisfies northwest corner. Hence λ satisfies the relation (42), where g is ~~b~~ given by (43). But there is just one measure satisfying this relation, so λ is unique. |||[16]

We now show that this result, establishing the existence and uniqueness of a weight-falloff measure on the plane, extends in part to the original problem on $S \times Q$.

Theorem: (uniqueness theorem) Let (S, Σ_s, α) and (Q, Σ_q, β) be measure spaces, and $h: S \to$ reals, $w: Q \to$ reals measurable functions such that α is finite from below and β finite from above (with respect to h, w, respectively). Let $\alpha\{s \mid h(s) = x\} = 0$ for all real numbers x, and let Σ_q be the class of all sets of the form $\{q \mid w(q) \in E\}$, E ranging over the Borel field on the real line.

Then there is at most one measure ν on $(S \times Q, \Sigma_s \times \Sigma_q)$, with marginals α, β, satisfying the (measurable) weight-falloff condition.

Proof: Consider the class, R, of sets of the form

$$F \times \{q \mid x < w(q) \leq y\}, \qquad (8.5.46)\ ~~(46)~~$$

where $F \in \Sigma_s$ and x, y are real numbers.

Class $\mathcal{R}$ generates the ~~sigma~~ σ-field $\Sigma_s \times \Sigma_q$. To show this, it suffices to prove that all measurable rectangles belong to Σ, the ~~sigma~~ σ-field with universe set $S \times Q$ generated by $\mathcal{R}$. Consider the class of Borel sets E on the real line having the property that $\{q|w(q) \in E\}$ belongs to the ~~sigma~~ σ-field on Q generated by the sets $\{q|x < w(q) \leq y\}$, x, y real. This class is closed under countable unions and complements; it also includes all half-open intervals $\{z|x < z \leq y\}$, x, y real. But the latter generate the Borel field; hence the sets $\{q|x < w(q) \leq y\}$ generate the ~~sigma~~ σ-field of sets $\{q|w(q) \in E\}$, E ranging over all real Borel sets. By assumption, this ~~sigma~~ σ-field is Σ_q. Hence Σ owns all measurable rectangles, so that $\Sigma = \Sigma_s \times \Sigma_q$.

Next, $\mathcal{R}$ is a semi-~~sigma~~ σ-ring (p. ~~page~~ 000). Indeed, $\emptyset \in \mathcal{R}$, the intersection of two $\mathcal{R}$-sets is an $\mathcal{R}$-set, and the difference of two $\mathcal{R}$-sets can be expressed as the union of three disjoint $\mathcal{R}$-sets. (In fact, $\mathcal{R}$ is a semi-ring.)

~~Now~~ let ν_1, ν_2 be two measures with marginals α, β, satisfying (measurable) weight-falloff. We ~~will~~ prove that ν_1 and ν_2 must coincide on $\mathcal{R}$, and that ~~there~~ exist a countable number of $\mathcal{R}$-sets whose union is $S \times Q$, such that ν_1 and ν_2 are finite on each. The basic extension theorem (p. ~~page~~ 000) then guarantees that ν_1 and ν_2 are equal, and we are through.

First, the $\mathcal{R}$-sets $S \times \{q|n < w(q) \leq n + 1\}$, $n = 0, \pm 1, \pm 2, \ldots$, cover $S \times Q$, and ν_1 and ν_2 must be finite on each, since their common right marginal, β, is finite from above. This proves half ~~of~~ the statement in the paragraph above.

18 It remains to show that ν_1, ν_2 coincide on R. Since they satisfy measurable weight-falloff, the measures, λ_1, λ_2, induced from them on the plane by the mapping $(s,q) \to (h(s), w(q))$ must satisfy weight-falloff, and must have as marginals the measures on the real line induced from α and β. These marginals are finite from below and above, respectively, so that ~~there~~ exists exactly one weight-falloff measure, λ, having them as marginals. Thus $\lambda = \lambda_1 = \lambda_2$. λ is, in fact, the northwest corner measure with these marginals. It follows that

break at = if necessary

$$\nu_i\Big\{(s,q)\,\Big|\,h(s) < x',\ w(q) > x\Big\} = \min\Big[\alpha\{s|h(s) < x'\},\ \beta\{q|w(q) > x\}\Big] \qquad (8.5.47)\ \cancel{(47)}$$

12 pt braces, brackets + vertical bar.

is true for all real numbers, x, x', for $i = 1,2$. For this is merely the northwest corner condition (35) expressed in terms of ν_i, α and β. (47) also holds for infinite x or x', as may be verified by direct substitution. x, x'

Next, for any number k between 0 and $\alpha(S)$, inclusive, ~~there~~ exists an extended real number y' such that

$$\alpha\{s|h(s) < y'\} = k. \qquad (8.5.48)\ \cancel{(48)}$$

To see this, take the supremum of the numbers y' for which the left side of (48) does not exceed k. For this value we have

$$\alpha\{s|h(s) < y'\} \le k \le \alpha\{s|h(s) \le y'\}, \qquad (8.5.49)\ \cancel{(49)}$$

from the "continuity" of measures. But, by assumption, the left and right terms in (49) are equal, hence (48) follows.

Now choose any member of $\check{R}$. Given the numbers x and y in (46) choose x' and y' to satisfy

$$\alpha\{s \mid h(s) < x'\} = \beta\{q \mid w(q) > x\}, \qquad \cancel{(50)}\ (8.5.50)$$

$$\alpha\{s \mid h(s) < y'\} = \beta\{q \mid w(q) > y\}. \qquad \cancel{(51)}\ (8.5.51)$$

Such x' and y' exist, by (48), since the right-hand terms in (50) and (51) lie between 0 and $\beta(Q) = \nu_i(S \times Q) = \alpha(S)$.

From (47) and (50) we obtain

$$\nu_i\{(s,q) \mid h(s) < x',\ w(q) > x\} = \alpha\{s \mid h(s) < x'\} = \beta\{q \mid w(q) > x\}, \qquad \cancel{(52)}\ (8.5.52)$$

for $i = 1,2$. The common value in (52) is finite, because x is finite, and β is finite from above. It follows that

$$\nu_i\{(s,q) \mid h(s) \geq x',\ w(q) > x\} = 0 = \nu_i\{(s,q) \mid h(s) < x',\ w(q) \leq x\}, \qquad \cancel{(53)}\ (8.5.53)$$

$i = 1,2$. To see this, note that the sum of the left-hand terms in (52) and (53) is $\nu_i\{(s,q) \mid w(q) > x\} = \beta\{q \mid w(q) > x\}$, and the first equality in (53) follows by subtraction. The second equality is proved similarly. The same argument applies with y, y' substituted for x, x', respectively, and we conclude that (53) remains true with these substitutions. Thus we have four equalities (53).

Now let $G = F \cap \{s|y' \leq h(s) < x'\}$. We then have

$$\nu_i\left[F \times \{q|x < w(q) \leq y\}\right] = \nu_i\left[G \times \{q|x < w(q) \leq y\}\right], \quad (8.5.54) \quad (54)$$

~~for~~ $i = 1,2$. For, the set of points (s,q) belonging to the left-, but not the right~~-hand~~ set in (54) is contained in the union of two of the four sets of measure zero of (53), as ~~one~~ we can verif~~ies~~y. Finally, ~~one has~~ we have

$$\nu_i\left[G \times \{q|x < w(q) \leq y\}\right] = \nu_i(G \times Q) = \alpha(G), \quad (8.5.55) \quad ~~(55)~~$$

$i = 1,2$. For the set of points (s,q) belonging to the middle, but not to the left~~-hand~~ set in (55) is contained in the union of the other two sets of measure zero of (53), as ~~one~~ we can verif~~ies~~y.

Equations (54) and (55) show that ν_1 and ν_2 coincide on all $\mathcal{R}$-sets. Hence they are identical. ~~|||~~ $\square$

Note that the assumptions imposed on the two component spaces are quite different, unlike all the other theorems of this section. The condition that $\alpha\{s|h(s) = x\} = 0$ states that the measure on the real line induced by h from α is non-atomic, while the condition that Σ_q is all sets of the form $\{q|w(q) \in E\}$ states that Σ_q is the ~~sigma~~ σ-field inversely induced by w from the real Borel field. From the symmetry of the allotment-assignment problem in S and Q, it is clear that these conditions could have been interchanged (making induced β nonatomic and Σ_s inversely induced) without invalidating the conclusion. But the form in which the theorem is stated is the one ~~which~~ that applies neatly to realistic Thünen systems.

Neither of these assumptions can be dropped without invalidating the conclusion. This will be illustrated later with counterexamples from simple Thünen systems.

Next we have an existence theorem for the original allotment-assignment problem similar to the one proved above for the transformed problem. A function is said to be <u>semi-continuous</u> iff it is either upper or lower semicontinuous (or both, i.e., continuous).

<u>Theorem</u>: Let (S,Σ_s,α) and (Q,Σ_q,β) be measure spaces, with $\alpha(S) = \beta(Q) < \infty$. Let Σ_s and Σ_q be the Borel fields of topologies T_s and T_q, respectively, these making S and Q Borel subsets of topologically complete and separable spaces. Let the functions $h{:}S \to$ reals and $w{:}Q \to$ reals be semi-continuous.

Then ~~there~~ exists a measure ν° on $(S \times Q,\ \Sigma_s \times \Sigma_q)$ with marginals α and β, which satisfies the (measurable) weight-falloff conditions (with respect to h, w).

<u>Proof</u>: First, let h and w be lower semi-continuous. Let $f{:}\text{reals}^2 \to$ reals be a function ~~which~~ that has positive cross-differences, and ~~which~~ is bounded, continuous, and increasing in each argument. An example is

$$f(x,y) = (1 + e^{-x})^{-1}(1 + e^{-y})^{-1} \; . \qquad \text{(8.5.56)}$$

We now show that the composite function $f(h(\cdot),w(\cdot)){:}S \times Q \to$ reals is lower semi-continuous. Let (s_o,q_o) belong to the set

$$\{(s,q)\,|\,f(h(s),\ w(q)) > z\},$$ (8.5.57) ~~(57)~~

z being a real number. Since f ~~being~~ is continuous, the set

$$\{(x,y)\,|\,f(x,y) > z\}$$ (8.5.58) ~~(58)~~

So 4 is open in the plane. Point $(h(s_o),\ w(q_o))$ belongs to (58), hence there is a point (x_o,y_o) southwest of $(h(s_o),\ w(q_o))$ ~~which~~ that belongs to (58). Consider the following subset of $S \times Q$:

$$\{s\,|\,h(s) > x_o\} \times \{q\,|\,w(q) > y_o\}.$$ (8.5.59) ~~(59)~~

(59) is open, by the lower semi-continuity of h and w. Point (s_o,q_o) belongs to (59) by construction. Finally, (59) is contained in (57), since f is increasing in its arguments. Hence (57) is an open set for any z, so that $f(h(\cdot),\ w(\cdot))$ is indeed lower semi-continuous.

It is also a bounded function, and these properties, ~~together~~ with the other premises, imply that the allotment-assignment problem (4), ~~(5),~~ (6) has a best solution $\nu°$ (Section 7.4). Since f has positive cross-differences, this $\nu°$ satisfies the measurable weight-falloff condition.

For the remaining three cases, replace f by f', where

displayed line for line

$f'(x,y) = f(-x,-y)$ if h, w are both upper semi-continuous;

$f'(x,y) = -f(-x,y)$ if h is lower and ~~f~~ w upper semi-continuous;

$f'(x,y) = -f(x,-y)$ if h is upper and ~~f~~ w lower semi-continuous,

f being given by (56).

In all cases f' remains bounded, continuous, with positive cross-differences (because the number of sign changes is even). And in all cases $f'(h(\cdot),\ w(\cdot))$ remains lower semi-continuous. This is clear for $f(-h(\cdot), -w(\cdot))$, since negation converts upper to lower semi-continuity. For $-f(-h(\cdot), w(\cdot))$ and $-f(h(\cdot), -w(\cdot))$, the two argument-functions of f are now upper semi-continuous. In this case, reasoning similar to that above shows that the composite function $f(\ldots)$ is upper semi-continuous. (Reverse the inequality signs in (57) & (58), ~~and (59),~~ and take (x_o, y_o) northeast rather than southwest). Hence $f'(\ldots) = -f(\ldots)$ is indeed lower semi-continuous.

As above, a measurable weight-falloff measure ν° then exists in all cases. □

Unlike the situation in the transformed problem, the ν° of this theorem need not be unique. A trivial example of this is where h or w is a constant and there are at least two feasible measures. For here every feasible measure ν satisfies both weight-falloff conditions vacuously.

We have obtained conditions under which an optimal solution must satisfy a weight-falloff condition. Our next result is a converse, indicating conditions under which a weight-falloff measure is optimal.

Theorem: Let (S,Σ_s,α) and (Q,Σ_q,β) be bounded measure spaces. Let the functions $h{:}S \to$ reals and $w{:}Q \to$ reals be measurable, and let $f{:}\text{reals}^2 \to$ reals be bounded, lower semi-continuous, with non-negative cross-differences. Let measure ν° on $(S \times Q, \Sigma_s \times \Sigma_q)$ have marginals α and β and satisfy the (measurable) weight-falloff condition.

Then ν^o is best for the allotment-assignment problem of minimizing

$13\frac{1}{2}\,\frac{1}{4}\,\frac{1}{8}$

$$\int_{S\times Q} f(h(s),w(q))\,\nu(ds,dq) \qquad \text{(8.5.60)} \; \cancel{(60)}$$

over measures ν with marginals α and β.

Proof: First ~~let us~~ assume that f has positive cross-differences. Consider the allotment-assignment problem on the plane induced from the given problem. Since f is bounded lower semi-continuous, α and β are bounded, and $\alpha(S) = \nu^o(S \times Q) = \beta(Q)$, ~~there exists~~ a best solution λ^{oo} exists to this transformed problem (cf. ~~page~~ Section 7.4). And λ^{oo} must satisfy the weight-falloff condition, since f has positive cross-differences.

Since ν^o satisfies weight-falloff, the measure λ^o on the plane induced from it by the mapping $(s,q) \to (h(s),w(q))$ must satisfy weight-falloff. λ^o is also feasible for the transformed problem. But there is only one weight-falloff measure feasible for the transformed problem, since α and β are bounded. Hence $\lambda^o = \lambda^{oo}$. λ^o is therefore unsurpassed for the transformed problem, implying that ν^o is unsurpassed for the original. But (60) is well-defined and finite for all feasible ν, so "unsurpassed" coincides with "best". This proves the theorem for the special case of positive cross-differences.

Now let f have merely non-negative cross-differences. Choose a function g:reals$^2 \to$ reals ~~which~~ that is bounded, lower semi-continuous, and has positive cross-differences (such as

(56)). Then, for any positive real number ε, the function $f + \varepsilon g$ has the same properties as g. Consider the perturbed allotment-assignment problem in which f in (60) is replaced by $f + \varepsilon g$. By the results just proved, ν° is best for this problem. Hence

$$\int_{S\times Q} (f + \varepsilon g)(h(s),w(q))\,\nu^\circ(ds,dq)$$

$$\leq \int_{S\times Q} (f + \varepsilon g)(h(s),w(q))\,\nu(ds,dq)$$

break at $\leq$ if necessary

for any other feasible measure ν and any real $\varepsilon > 0$. Now let ε go to zero. By the dominated convergence theorem, the limit of the integral on each side is the integral of the limit. Hence ν° remains best when $\varepsilon = 0$. ~~This completes the proof.~~ $\square$

We ~~shall~~ later show that this theorem can be strengthened to some extent, viz. ~~Namely,~~ the premise that f is lower semicontinuous can be dropped. But the method of proof just used is quite instructive and completely different from the method to be used below, which involves the construction of a potential.

The resulting theory is ~~a~~ fairly satisfactory ~~one~~, and the conditions under which it holds are ~~, for the most part,~~ not too onerous. The boundedness of f, however, is a nuisance; e.g., ~~For example,~~ the product function, $f(x,y) = xy$ (which is the original form in which transport cost presented itself) is not bounded on the plane. This limitation is easily remedied if the ideal distance and weight functions h and w have bounded ranges. For then we can take the domain of the transformed problem to be – not the entire plane – but a rectangle with bounded intervals as sides. On such a set the product function (and most other functions of interest) will be bounded, and the preceding theorem can be applied.

Insight into the distinction between positive and nonnegative cross-differences can be gained by contemplating the case where $\underline{f}$ has zero cross-differences, ~~that is~~ i.e., where

$$f(x_1,y_1) + f(x_2,y_2) = f(x_1,y_2) + f(x_2,y_1), \qquad (8.5.61)$$

for all numbers x_1, x_2, y_1, y_2. Equation (61) holds iff $\underline{f}$ can be written as the sum of separate x- and y-functions:

$$f(x,y) = f_1(x) + f_2(y). \qquad (8.5.62)$$

(Proof: If (62) holds, then (61) is verified by substitution. Conversely, choose an arbitrary y_0, and define f_1, f_2 by: $f_1(x) = f(x,y_0)$, $f_2(y) = f(x,y) - f(x,y_0)$. For the definition of f_2 to be sound, the expression $f(x,y) - f(x,y_0)$ must not depend on x. But this is guaranteed by (61), and (62) follows at once.) But if (62) holds (and f, α, β are all bounded), then the objective function (60) is equal to

$$\int_S (f_1 \circ h)\,d\alpha + \int_Q (f_2 \circ w)\,d\beta,$$

by the induced integrals theorem. ("$\circ$" signifies the composition of functions). Thus transport cost depends only on the marginals, α and β, of ν. Since all feasible ν have the same marginals, they are <u>all</u> best solutions. Thus, while positive cross-differences restrict best solutions to the weight-falloff measures, non-negative cross-differences may allow others.

B

<u>Potentials</u>

We now turn out attention to the construction of <u>potentials</u>. This is of interest not only for the further insights it furnishes concerning the optimality properties of weight-falloff measures, but also because potentials have direct intuitive interpretations: as <u>land values</u> and as "gross profits" on land uses.

Let measure ν° on $(S \times Q, \Sigma_s \times \Sigma_q)$ have left- and right-marginals α and β, so that it is feasible for the allotment-assignment problem. Recall that a pair of measurable functions, $p: S \to$ reals and $k: Q \to$ reals is a <u>measure potential</u> for ν° (in the wide sense) iff

$$k(q) - p(s) \leq f(h(s), w(q)) \qquad \text{~~(63)~~ (8.5.63)}$$

for all $s \in S$, $q \in Q$, and

12 pt fences & vertical bar

$$\nu^\circ\left\{(s,q) \,\middle|\, k(q) - p(s) < f(h(s), w(q))\right\} = 0.^{17} \qquad \text{~~(64)~~ (8.5.64)}$$

~~New~~ Now furnish S and Q with topologies $\mathcal{T}_s$ and $\mathcal{T}_q$, respectively. The pair of measurable functions (p,k) is said to be a <u>topological potential</u> for ν° (in the wide sense) iff (63) holds for all $s \in S$, $q \in Q$, and, if (s,q) is a point of support for ν°, then (63) holds with <u>equality</u>.

~~$k(q) - p(s) = f(h(s), w(q))$.~~

We shall make essential use of the transformation of the allotment-assignment problem into the plane. The transformed problem is itself a special case of the allotment-assignment problem, and we may therefore contemplate potentials for *its* feasible solutions, λ. Our first task will be the following. Let ν be feasible for the original problem, and λ the measure induced from ν; λ is feasible for the transformed problem. What relations then hold between the properties of there being potentials (measure- or topological-) for λ and for ν?

First, consider one preliminary. We ~~shall~~ take the transformed problem to be defined, not necessarily on the whole plane, but on a rectangular subset of the plane, $X \times Y$. The exact definition will be given later; ~~and~~ for the present we need merely assume that X and Y contain the ranges of h and w, respectively. A potential for the transformed problem is then a pair of measurable functions, $p:X \to$ reals, $k:Y \to$ reals, satisfying one or the other of the definitions above. Thus $X \times Y$ is the domain of f and the universe set of feasible measures λ.

Theorem: Let ν° be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$ with left and right marginals α and β. Let X and Y be measurable subsets of the real line; and let $h:S \to X$, $w:Q \to Y$, $f:X \times Y \to$ reals, $p:X \to$ reals, and $k:Y \to$ reals be measurable functions. Let λ° be the measure on $X \times Y$ induced from ν° by the mapping $(s,q) \to (h(s), w(q))$, and consider the following three conditions:

centered degree sign $\overset{\vee}{p}\,{}^{\circ}_{n}\,h$ etc

align (:) right

(i)	(p,k) is a topological potential for λ°;
(ii)	(p,k) is a measure potential for λ°;
(iii)	($p\circ h$, $k\circ w$) is a measure potential for ν°.[18]

Then condition (i) implies condition (ii), which in turn implies condition (iii).

Proof: The usual topology on the plane, or any subset of the plane, has the strong Lindelöf property. This insures that any topological potential is a measure potential (Section 7.5). Thus (i) implies (ii).

Let condition (ii) be valid. Then

$$k(y) - p(x) \leq f(x,y)$$

for all $x \in X$, $y \in Y$. Letting $x = h(s)$, $y = w(q)$, we verify (63) for the pair of functions ($p\circ h$, $k\circ w$). Also we have

$$\nu^\circ\left\{(s,q)\,\middle|\,k(w(q)) - p(h(s)) < f(h(s),\, w(q))\right\}$$
$$= \lambda^\circ\left\{(x,y)\,\middle|\,k(y) - p(x) < f(x,y)\right\} = 0. \qquad (65)$$

12 pt braces + vertical bar (8.5.65)

The right equality is from condition (ii), the left from the fact that the argument of ν° is the inverse image of the argument of λ° under the mapping $(s,q) \to (h(s), w(q))$. Condition (65) yields (64) for the pair of functions ($p\circ h$, $k\circ w$). Thus (iii) is valid. ∎

start 27-7

This theorem is silent about topological potentials for ν°. Indeed this concept is not even defined, since nothing is said about any topologies on S or Q.

Our plan of action is to construct a topological potential for the weight-falloff measure λ°. If λ° is induced from a measure ν° feasible for the original allotment-assignment problem, the preceding theorem yields a measure potential for ν°. From this we ~~one~~ can make inferences concerning the optimality of ν°.

In constructing a potential for λ° we could make use of the theory developed for the transportation problem. Instead, ~~however,~~ we use a special procedure that ~~which~~ utilizes the distinctive properties of weight-falloff and non-negative cross-differences. This not only allows us to weaken the assumptions needed, but the procedure is of interest in itself and has intuitive appeal.

We begin with an observation. A measure λ on the plane has all its mass concentrated on its support. That is, if E is the support of λ, then the complement of E has measure zero. (E and its complement are Borel sets, since E is closed). The proof of this rests on the strong Lindelöf property of the usual topology of the plane, For every point of the complement of E has a measurable neighborhood of measure zero. A countable subcollection of these neighborhoods covers the complement of E, which therefore has measure zero.

Given measure λ on the plane, we ~~shall~~ restrict it to a rectangle $X \times Y$ as follows. X consists of all numbers x having the property that either there is a number y such that (x,y) supports λ, or x is between two such numbers, x_1 and x_2.

Y is defined analogously, the rôles of x and y being interchanged. X is an interval, which may be the entire real line, or may be bounded below, or above, or both. If bounded, the endpoints may or may not be included in X. The same remarks apply to Y. The support E of λ is contained in $X \times Y$. Hence, by the preceding argument, we are throwing away a set of measure zero. Finally, note that X and Y are empty iff λ has empty support. By the preceding argument this occurs iff $\lambda = 0$. We exclude this trivial case by assumption. Call $X \times Y$ the <u>support rectangle</u>.[19]

Now let λ be a measure on $X \times Y$ satisfying the weight-falloff condition. With the point of support (x,y) associate the value $x-y$. No two points of support have the same value, for if this were true of (x_i,y_i), $i = 1,2$, we would have $x_1 - x_2 = y_1 - y_2 \neq 0$; the points would thus stand in a southwest-northeast relation, contradicting weight-falloff. The valuation thus determines a complete anti-symmetric ordering of the points of support. We have, in fact, a "Maginot line" of points of support strung across the plane, running from northwest to southeast (possibly including vertical, north-to-south, stretches, and/or horizontal, west-to-east, stretches).

This line may have gaps in it. A <u>gap</u> is defined as a pair of distinct points of support, (x_1,y_1), (x_2,y_2), with no other points of support "between" them in the ordering. Wherever such a gap exists, connect the two points constituting it by a <u>straight line-segment</u>. The union of the original support and all these line-segments is called the line of support, L, of λ°. It is easy to prove the following facts about the line of support. No two points of it stand in a southwest-northeast relation. For every $x \in X$, ~~there~~ exists a $y \in Y$ such that $(x,y) \in L$. For given $x \in X$, there is either a unique $y \in Y$ such that $(x,y) \in L$, or a closed interval (possibly unbounded) of such y's. Similar statements apply with x and y interchanged. The line of support is contained in the rectangle of support.

Having furnished the line of support, the measure λ° has completed its rôle in the construction of a topological potential for itself, and attention now passes to the cost function $f{:}X \times Y \rightarrow$ reals. From f we ~~shall~~ construct a function $p{:}X \rightarrow$ reals, which turns out to be (under certain conditions) the left half of a topological potential for λ°.

27 Function

$p(x)$ is defined as follows. Choose a fixed point $x_0 \in X$, and define $p(x_0) = 0$. For $x \neq x_0$, take a sequence (x_i, y_i), $i = 0, \ldots, n$, of points on the line of support, with $x_n = x$. This sequence is to be monotone, ~~that is~~ i.e., either strictly increasing in the value $x_i - y_i$ (for $x > x_0$), or strictly decreasing in $x_i - y_i$ (for $x < x_0$); n is any positive integer. With this sequence associate the real number

$$(x_0 y_1 - x_1 y_1) + (x_1 y_2 - x_2 y_2) + \ldots + (x_{n-1} y_n - x_n y_n), \qquad (66)$$

(8.5.66)

where we use the abbreviation "xy" for $f(x,y)$ here and below. $p(x)$ is then defined as the infimum of (66) over all such monotone sequences from x_0 to x.

#

Lemma: Let measurable $f: X \times Y \to$ reals have non-negative cross-differences. Then p is finite, ~~real-valued~~, measurable, and

$$p(x') + f(x',y') \le p(x) + f(x,y') \qquad (8.5.67) \quad \cancel{(62)}\ 7$$

for any $x, x' \in X$, $y' \in Y$ such that $(x',y') \in L$, the line of support.

#

Proof: Let (x_i, y_i), $i = 0, \ldots, n$, be a monotone sequence of L-points for $x = x_n$. Then

$$68 \quad (x_{i-1}y_0 - x_i y_0) \le (x_{i-1}y_i - x_i y_i) \le (x_{i-1}y_n - x_i y_n),$$

$i = 1, \ldots, n$. This follows at once from non-negative cross-differences and the monotonicity of the sequence. Adding these inequalities over i, we obtain

$$(x_0 y_0 - x_n y_0) \le z \le (x_0 y_n - x_n y_n), \qquad (8.5.68) \quad (68)$$

where z is of the form (66). From this we obtain

$$x_0 y_0 - x y_0 \le p(x) \le x_0 y - x y, \qquad (8.5.69) \quad \cancel{(69)}$$

where y_0 and y are any two numbers such that $(x_0, y_0) \in L$ and $(x,y) \in L$. The right inequality in (69) follows from the right of (68) by taking the infimum of z over all permissible sequences with fixed $(x,y) = (x_n, y_n) \in L$ (n any positive integer). The left inequality in (69) follows from the left of (68) by taking the infimum of z over all permissible sequences with fixed y_0, on noting that this does not restrict the range of values (66). Inequality (69) shows that p is indeed finite ~~real-valued~~.

$x_0 y_0$

$|(x_0, y_0)$ zero

Inequality (67) follows at once for two special cases besides the obvious $x = x'$; for $x' = x_0$, from the left of (69), and for $x = x_0$ from the right of (69). This leaves the case where x_0, x, and x' are all distinct. There are three subcases, depending on which of these numbers is between the other two.

(i) If x_0 is between x and x', we have

$$p(x') \leq x_0y' - x'y' \qquad \text{(8.5.70)} \;\; \cancel{(70)}$$

~~and~~

$$p(x) \geq x_0y_0 - xy_0 \geq x_0y' - xy'. \qquad \text{(8.5.71)} \;\; \cancel{(71)}$$

Inequality (70) and the left inequality in (71) arise from (69); the right of (71) arises from nonnegative cross-differences. (70) and (71) together yield (67).

(ii) If x is between x_0 and x', we consider monotone sequences (x_i, y_i), $i = 0, \ldots, n$, of L-points, with $(x_n, y_n) = (x', y')$, which include x among the x_i's, say $x = x_j$. Separate the corresponding expression (66) into two sums: the first $2j$ terms (ending with $-x_jy_j$), and the last $2n-2j$ terms (beginning with x_jy_{j+1}). This latter sum does not exceed $x_jy_n - x_ny_n$. (The argument for this is the same as that leading to the right-hand inequality of (68), except that we begin with x_j, not x_0.) Hence we have

$$p(x') \leq (x_0y_1 - x_1y_1) + \ldots + (x_{j-1}y_j - x_jy_j) + x_jy_n - x_ny_n. \qquad \text{(8.5.72)} \;\; (72)$$

fl l 279

Now (72) holds for all monotone sequences $(x_0, y_0), \ldots, (x_j, y_j)$ of L-points for which $x_j = x$. Taking the infimum over the corresponding sums (66), we obtain

$$p(x') \leq p(x) + x_j y_n - x_n y_n,$$

which is the same as (67).

18 ¶ (iii) Finally, if x' is between x_0 and x, we note first that removing a point (x_j, y_j) (where $0 < j < n$) does not decrease the corresponding sum (66). For, the change in (66) is

minor here

$$(x_{j-1}y_{j+1} - x_{j+1}y_{j+1}) - (x_{j-1}y_j - x_j y_j) - (x_j y_{j+1} - x_{j+1}y_{j+1}),$$

which is ≥ 0, by non-negative cross-differences and the monotonicity of the sequence (x_i, y_i), $i = 0, \ldots, n$, of L-points.

Now, for any $\varepsilon > 0$, we can find a sequence of L-points such that the sum (66) does not exceed $p(x) + \varepsilon$. If (x', y') is not among these points, slip it into the sequence so as to preserve monotonicity: say $(x', y') = (x_j, y_j)$ after relabeling. By the observation just made, this insertion cannot increase (66), so ~~that~~ it remains $\leq p(x) + \varepsilon$. Now separate (66) into two sums as above. The first ~~sum~~ (which ends with the term $-x_j y_j$) is at least as large as $p(x')$. The second ~~sum~~ (which begins with $x_j y_{j+1}$) is at least as large as $x_j y_j - x_n y_j$. Use the ~~(S~~same argument ~~as~~ that leads to the left inequality of (68) with j in place of 0.~~)~~ Hence

$$p(x) + \varepsilon \geq p(x') + x_j y_j - x_n y_j.$$

That is,

$$p(x) + \varepsilon \geq p(x') + x'y' - xy'.$$

Since $\varepsilon > 0$ is arbitrary, we again obtain (67).

It remains only to prove that p is measurable. We show that p restricted to the bounded interval from x_0 to x° is measurable.[20] Since x° is arbitrary, this implies that p itself is measurable.

For each $m = 1,2,\ldots$, let (x_{mi}, y_{mi}), $i = 0,\ldots,n_m$, be a monotone sequence of L-points, with $x_{m0} = x_0$ and $x_{mn_m} = x^{\circ}$ for all m, such that

$$p(x^{\circ}) + \frac{1}{m} \geq (x_{m0}y_{m1} - x_{m1}y_{m1}) + \ldots + (x_{m,n_m-1}y_{mn_m} - x_{mn_m}y_{mn_m}). \qquad (73)$$

Now let (x_i, y_i), $i = 0,\ldots,n$, be a monotone sequence of L-points, with $x_n = x^{\circ}$. In terms of this sequence, we define the function p' on the (x_0, x°) interval as follows. First, for each number x in the interval, choose a $y(x)$ for which $(x, y(x)) \in L$. Then

$$p'(x) = (x_0y_1 - x_1y_1) + \ldots + (x_{j-1}y_j - x_jy_j) + (x_jy(x) - xy(x)).$$

Here j is such that $(x, y(x))$ lies between (x_j, y_j) and (x_{j+1}, y_{j+1}) (possibly coinciding with the latter) in the natural ordering of L-points.

p' is a measurable function. For, $x_jy(x) = f(x_j, y(x))$ and $xy(x) = f(x, y(x))$ are both measurable, since f itself is measurable, and $y(x)$, being monotone, is measurable. Thus p'

is measurable on the interval from x_j to x_{j+1}, hence on the whole interval (x_0, x°).

Now let p_m be the p'-function defined in terms of the sequence (x_{mi}, y_{mi}), $i = 0,\ldots,n_m$. We claim that, for all x in the interval (x_0, x°),

$$p_m(x) \geq p(x) \geq p_m(x) - 1/m. \tag{8.5.74}$$

The left inequality in (74) is immediate. To prove the right inequality, insert the point $(x, y(x))$ in the sequence (x_{mi}, y_{mi}), $i = 0,\ldots,n_m$, and make the corresponding change on the right side of (73). Since this side does not increase, (73) remains valid. On the right side, the sum of terms up to $-xy(x)$ is $p_m(x)$, so we have

$$p(x^\circ) + 1/m \geq p_m(x) + z. \tag{8.5.75}$$

Here z is the sum of the remaining terms; it begins with xy_{mj} for some j. Now let (x_i, y_i), $i = 0,\ldots,n$, be a monotone sequence of L-points, with $x_n = x$. We have

$$p(x^\circ) \leq (x_0y_1 - x_1y_1)+\ldots+(x_{n-1}y_n - x_ny_n) + z.$$

27-10

Taking the infimum over all such sequences, we obtain

$$p(x^\circ) \leq p(x) + z. \tag{8.5.76}$$

(75) and (76) together yield the right inequality of (74).

But (74) implies that $p(x)$ is the limit of $p_m(x)$ as $m \to \infty$, for all x in the interval (x_0, x°). As the limit of a sequence of measurable functions, p itself (restricted to (x_0, x°)) is measurable. ~~The proof is complete.~~

This lemma implies that p, as defined by (66) is a <u>left half-potential</u> for λ° (see (7.5.8) ~~(8) of 7.5~~). For, if (x', y') supports λ°, then $(x', y') \in L$, and (67) is ~~then~~ the half-potential condition for p.

CONT. 10

Theorem: Let α, β be σ ~~sigma~~-finite measures on the real intervals X, Y, respectively, and let measurable $f: X \times Y \to$ reals have non-negative cross-differences. Let λ° satisfy the weight-falloff condition and be feasible for the allotment-assignment problem of minimizing $\int f\, d\lambda$, subject to the constraints $\lambda' = \alpha$, $\lambda'' = \beta$. Let $X \times Y$ be the rectangle of support for λ°.

Then ~~there~~ exists a (p,k) that ~~which~~ is both a topological and measure potential for λ° (wide sense).

Proof: Any topological potential is a measure potential here, hence we need only construct the former. Construct L, the line of support for λ°, and then construct $p: X \to$ reals according to (66). Now define the function $k: Y \to$ reals by

$$k(y) = \inf\{p(x) + f(x,y)\}, \qquad (8.5.77)\ ~~(77)~~$$

the infimum taken over all $x \in X$. Function k is ~~indeed~~ finite; ~~real valued.~~ for, by (67), the infimum is attained at any x such that $(x, y) \in L$, and such an x exists for each y.

It follows at once from (77) that

$$k(y) - p(x) \leq f(x,y) \qquad (8.5.78)$$

for all $x \in X$, $y \in Y$. Furthermore, (78) is satisfied with equality if (x,y) supports λ^o. For in this case $(x,y) \in L$, and the infimum of (77) is attained at this x.

Next we show that k is measurable. For each $y \in Y$, choose an $x(y)$ such that $(x(y), y) \in L$. Then

$$k(y) = p(x(y)) + f(x(y), y). \qquad (8.5.79)$$

$x(y)$ is ~~a~~ monotone function, hence measurable. Also f and p are ~~also~~ measurable, and (79) then shows that k is measurable.

Thus the pair (p,k) is a topological potential for λ^o. |||

The properties of p and k, other than measurability, are unspecified. However, if one makes further assumptions about f one can say more.

Theorem: Assume the premises of the preceding theorem, and construct the potential (p,k) according to its proof. Then:

(i) If f is bounded, p and k are *bounded*.

(ii) Consider f as a family of functions $f(\cdot,y): X \to$ reals indexed by $y \in Y$. If these functions are all strictly (increasing, decreasing), ~~then~~ p is *strictly* (*decreasing*, *increasing*), respectively.

(iii) Consider f as a family of functions $f(x,\cdot): Y \to$ reals indexed by $x \in X$. If these functions are all strictly (increasing, decreasing), ~~then~~ k is *strictly* (*increasing*, decreasing), respectively.

(iv) If the functions $f(\cdot,y)$, $y \in Y$, are all concave, then p is convex; if the functions $f(x,\cdot)$, $x \in X$, are all concave, then k is concave.

(v) If the functions $f(\cdot,y)$, $y \in Y$, are all upper semi-continuous, then p is lower semi-continuous; if the functions $f(x,\cdot)$, $x \in X$, are all upper semi-continuous, then k is upper semi-continuous.

(vi) If the family in (ii) is equicontinuous at $x_0 \in X$, then p is continuous at x_0; if the family in (iii) is equicontinuous at $y_0 \in Y$, then k is continuous at y_0.

(vii) If the family in ((ii), (iii)) is uniformly equicontinuous, then (p,k) is uniformly continuous, respectively.

Proof: (i) Boundedness of p follows from (69), and boundedness of k from that of p and f.

(ii) Let $f(\cdot,y)$ be strictly increasing for all $y \in Y$, and let $x,x' \in X$ satisfy $x < x'$. Choosing $y' \in Y$ so that $(x',y') \in L$, we obtain from (67):

$$p(x) - p(x') \geq f(x',y') - f(x,y') > 0,$$

so p is strictly decreasing. If $f(\cdot,y)$ is strictly decreasing, all $y \in Y$, let $x > x'$. The same argument yields $p(x) > p(x')$ again, so p is strictly increasing.

(iii) Since the infimum in (77) is attained at any x such that $(x,y) \in L$, we have

$$k(y') - k(y) \geq f(x',y') - f(x',y), \qquad \text{(8.5.80)} \; \cancel{(80)}$$

valid for all $x' \in X$, $y,y' \in Y$ such that $(x',y') \in L$. If now $f(x,\cdot)$ is strictly increasing, all $x \in X$, choose $y' > y$. Then (80) ~~then~~ implies that $k(y') > k(y)$, so k is strictly increasing. If $f(x,\cdot)$ is strictly decreasing, choose $y' < y$. (80) again yields $k(y') > k(y)$, so k is strictly decreasing.

(iv) For each $x \in X$ there is a $y \in Y$ such that $(x,y) \in L$. Since (77) is satisfied with equality at such a point (x,y), we obtain

$$p(x) = \sup\{k(y) - f(x,y)\}, \qquad \text{(8.5.81)} \; \cancel{(81)}$$

the supremum taken over all $y \in Y$. ~~Now,~~ for fixed y, $k(y) - f(\cdot,y)$ is convex. Hence p, as the supremum of a family of convex functions, is convex. Similarly, from (77), k is the infimum of a family of concave functions, hence is itself concave.

(v) For fixed $y \in Y$, $k(y) - f(\cdot,y)$ is lower semi-continuous. From (81), p is the supremum of a family of such functions, hence is itself lower semi-continuous. Similarly, from (77), k is the infimum of a family of upper semi-continuous functions, hence is itself upper semi-continuous.

(vi) and (vii): Treating p, k again as the supremum, infimum of a certain family of functions, repeat the arguments ~~already~~ given in Section 8.0 pages 000–00.

Under certain conditions, p and k can be expressed as <u>line integrals</u> along the line of support. These are defined as follows. Let g be a bounded measurable function whose domain is a subset of the plane containing the line of support, and let $x_1 \le x_2$, where $x_1, x_2 \in X$. Then we define

$$\oint_{x_1}^{x_2} g\,dx = \int_{(x_1,x_2)} g(x,y(x))\,dx. \qquad (8.5.82)\ \cancel{(82)}$$

On the right of (82), y(x) is any function such that $(x, y(x)) \in L$ for all $x \in X$. The integral is then over the open interval (x_1, x_2) with respect to Lebesgue measure. For this to be a *bona fide* definition, the value of the integral must not depend on the particular function y(x) chosen. That this is so may be seen as follows. For given $x \in X$, the set of numbers y such that $(x,y) \in L$ is either a singleton or an entire interval, the interiors of two such intervals being disjoint. Since each such interval has a rational number in its interior, there are at most a countable number of them. Hence, except for a countable number of x-values, y(x) is uniquely determined. But the Lebesgue measure of a countable set is zero, so ~~that~~ changes in y(x) on this set do not affect (82). Hence the definition is sound. This is called the integral of g along the line of support with respect to x, from x_1 to x_2.

It is also convenient to define this in the case when $x_1 > x_2$ by the rule

$$\oint_{x_1}^{x_2} g\,dx = -\oint_{x_2}^{x_1} g\,dx.$$

$\oint = -\oint$

The line integral of g with respect to y, from y_1 to y_2 ($y_1 < y_2$) is defined analogously, with $g(x(y),\ y)$ replacing $g(x,\ y(x))$ on the right of (82).

In the following theorem recall that $D_i f(x,y)$ represents the partial derivative of f at the point (x,y) with respect to its i-th argument, ($i = 1,2$).

Theorem: Let all the premises of the theorem above (page 000) be satisfied, and let the potential (p,k) be defined by (66) and (77). Let G be an open subset of the plane containing $X \times Y$, and let $f'{:}G \to$ reals coincide with f on $X \times Y$. Then

(i) if $D_1 f'(x,y)$ exists and is continuous on G, we have

$$p(x_2) - p(x_1) = \oint_{x_2}^{x_1} D_1 f(x,y)\,dx \qquad (8.5.83)$$

for all x_1, $x_2 \in X$; ~~and~~

(ii) if $D_2 f'(x,y)$ exists and is continuous on G, we have

$$k(y_2) - k(y_1) = \oint_{y_1}^{y_2} D_2 f(x,y)\,dy \qquad (8.5.84)$$

align () right

for all $y_1, y_2 \in Y$.

Proof: (i) Note first that the integrands $D_i f$ in (83) and (84) exist and coincide with $D_i f'$, (i = 1,2, respectively). To prove (83), we need to demonstrate only the special case where $x_2 = x_0$, x_0 being the special number used in the definition of p; for then (83) follows in general by subtraction. Thus we must show that

$$-p(x^o) = \int_{x_0}^{x^o} D_1 f(x,y)\,dx \qquad \text{(8.5.85)} \quad \cancel{(85)}$$

for all $x^o \in X$.

Take numbers y_0, y^o, such that $(x_0,y_0) \in L$ and $(x^o,y^o) \in L$, and consider the closed, bounded rectangle F having the points (x_0,y_0), (x^o,y^o), (x^o,y_0), and (x_0,y^o) as ~~at opposite~~ corners. By assumption, $D_1 f'$ is continuous on F; hence it is ~~in fact~~ bounded and uniformly continuous on F. Boundedness insures that the integral (85) is well-defined. By uniform continuity, for all m = 1,2,... there is a $\delta_m > 0$ such that, if $|y' - y''| < \delta_m$, then

$$|D_1 f(x,y') - D_1 f(x,y'')| \leq \frac{1}{m}, \qquad \text{(8.5.86)} \quad \cancel{(86)}$$

for all x between x_0 and x^o, all y', y" between y_0 and y^o.

Now take a sequence of points along the line of support from (x_0,y_0) to $(x_n,y_n) = (x^o,y^o)$, such that $|y_{i-1} - y_i| < \delta_m$ for i = 1,...,n. Then, for each i = 1,...,n,

subscript zero

do not break

$$\left|\int_{x_i}^{x_{i-1}} D_1 f(x,y_i)\,dx - \oint_{x_i}^{x_{i-1}} D_1 f(x,y)\,dx\right| \le \frac{1}{m}\left|x_i - x_{i-1}\right|$$

from (86), since, for each x between x_{i-1} and x_i, the $y(x)$ such that $(x,y(x)) \in L$ differs from y_i by less than δ_m. Now

$$\int_{x_i}^{x_{i-1}} D_1 f(x,y_i)\,dx = f(x_{i-1},y_i) - f(x_i,y_i). \qquad \text{(8.5.87)}\ \cancel{(87)}$$

The right of (87) is a typical pair of terms in the summation (66) for the sequence (x_i, y_i), $i = 0,\ldots,n$. It follows that the sum (66) differs from the line integral

$$\oint_{x^\circ}^{x_0} D_1 f(x,y)\,dx \qquad \text{(8.5.88)}\ \cancel{(88)}$$

by ~~less than~~ at most $\frac{1}{m}|x^\circ - x_0|$, if the differences between successive y_i's are all less than δ_m.

Now take a sequence, the m-th member of which is itself a sequence of points along L whose corresponding sum (66) is within $\frac{1}{m}$ of $p(x^\circ)$. Add to this m-th sequence, if necessary, sufficient extra points so that successive y_i's differ by less than δ_m. These additions do not increase (66), so ~~that~~ it remains within $\frac{1}{m}$ of $p(x^\circ)$; it is also within $\frac{1}{m}|x^\circ - x_0|$ of (88). Letting $m \to \infty$, we conclude that (88) equals $p(x^\circ)$. This is the same as (85), which implies (83).

(ii) To prove (84), we need yet another expression for k, one similar to (66). Specifically, we now show that

12 pt brackets

$$k(y^o) = \inf\left[(x_{n-1}y_n - x_{n-1}y_{n-1}) + \ldots + (x_1y_2 - x_1y_1) + x_0y_1\right], \qquad (89)$$

(8.5.89)

for all $y^o \in Y$ ("xy" abbreviates $f(x,y)$). Here the infimum is to be taken over all monotone sequences of points, (x_i, y_i), $i = 0,\ldots,n$, along the line of support such that $y_n = y^o$. (n can be any positive integer).

To prove (89), note that the bracketed expression ~~in (89)~~ is simply $z + x_ny_n$, where z is the sum (66). Since $k(y^o) = p(x_n) + x_ny_n \leq z + x_ny_n$, this proves (89) with the sign "$\leq$" substituted for "=". To prove the opposite inequality, choose a number x^o for which $(x^o,y^o) \in L$, ~~and~~ then a sequence of L-points such that (66) comes within ε of $p(x^o)$:

$$p(x^o) + \varepsilon \geq (x_0y_1 - x_1y_1) + \ldots + (x_{n-1}y_n - x_ny_n). \qquad (90)$$

(8.5.90)

Here $x_n = x^o$. We may also assume that $y_n = y^o$; for if not, insert the point (x^o,y^o) in the sequence. This does not disturb the validity of (90), but does make $(x_{n-1}y_n - x_ny_n) = 0$, since $x_{n-1} = x_n = x^o$. Hence we can keep or delete the last parenthetical term in (90), allowing us to make $-x^oy^o$ the last term. Adding x^oy^o to both sides, and letting $\varepsilon \to 0$, we obtain (89) with the opposite inequality. This proves (89).

Now choose a number y_0 such that $(x_0,y_0) \in L$. Note that $k(y_0) = x_0y_0$, since $p(x_0) = 0$. Hence

$$k(y^o) - k(y_0) = \inf\left[(x_{n-1}y_n - x_{n-1}y_{n-1}) + \ldots + (x_0y_1 - x_0y_0)\right].$$

12 pt brackets

pl

At this point we repeat the argument of the first half of this proof, interchanging the rôles of x and y, to conclude that

$$k(y^\circ) - k(y_0) = \int_{y_0}^{y^\circ} D_2 f(x,y)\, dy.$$

Letting $y^\circ = y_2$, and then $y^\circ = y_1$, and subtracting, we obtain (84). |||

Note, ~~by the way,~~ that in (83) x_1 is on top, while in (84) y_2 is on top. For the product function $f(x,y) = xy$, we have $D_1 f(x,y) = y$, $D_2 f(x,y) = x$, so that the line-integrals (83), (84) take on an especially simple form in this case.

We ~~Let us~~ now return to the problem of showing that weight-falloff measures are optimal for allotment-assignment.

D

Theorem: Let (S,Σ_s,α) and (Q,Σ_q,β) be bounded measure spaces, let $h:S \to$ reals and $w:Q \to$ reals be measurable, and let ν° be a measure on $(S \times Q,\ \Sigma_s \times \Sigma_q)$, with marginals α and β, satisfying the (measurable) weight-falloff condition (with respect to h, w). Let bounded measurable $f:\text{reals}^2 \to$ reals have non-negative cross-differences.

Then ν° is best for the allotment-assignment problem of minimizing

$$\int_{S\times Q} f(h(s),\ w(q))\,\nu(ds,dq) \qquad (8.5.91)$$

over measures ν with marginals α and β.

Proof: Let λ^o be the measure on the plane induced from ν^o by the mapping $(s,q) \to (h(s), w(q))$. Since ν^o satisfies measurable weight-falloff, so does λ^o. Let $X \times Y$ be the rectangle of support for λ^o. If $X \times Y = \emptyset$, then $\nu^o = 0$, so $\alpha = \beta = 0$ and the theorem is trivial. We may, therefore, assume that $X \times Y$ is not empty. Let $S_1 = \{s | h(s) \in X\}$ and $Q_1 = \{q | w(q) \in Y\}$. The complement of $X \times Y$ on the plane has λ^o-measure zero; hence $\nu^o((S \times Q) \setminus (S_1 \times Q_1)) = 0$, $\alpha(S \setminus S_1) = 0$, and $\beta(Q \setminus Q_1) = 0$.

Now consider the modified allotment-assignment problem in which α and h are restricted to S_1, β and w are restricted to Q_1, feasible measures ν have universe set $S_1 \times Q_1$, and the integral (91) is over $S_1 \times Q_1$. Any measure feasible for the original problem has the form $\nu \oplus 0$, where ν is feasible for the modified problem, and 0 is the zero measure on display don't break

$$((S \times Q) \setminus (S_1 \times Q_1)).$$

This establishes a 1-1 correspondence between the feasible solutions to these two problems, and the values of the objective function (91) for corresponding solutions are equal. Hence we need prove only that the restriction of ν^o to $S_1 \times Q_1$ is best for the modified problem.

~~Now~~ Let α_1, β_1, h_1, w_1 and ν_1^o denote the appropriate restrictions of these functions for the modified problem. Also let f_1, λ_1^o be the restrictions of f and λ^o to $X \times Y$. Since λ_1^o satisfies weight-falloff and f_1 has nonnegative cross-differences, ~~there~~ exists a measure potential, (p,k), for λ_1^o.

Hence $(p \circ h_1, k \circ w_1)$ is a measure-potential for ν_1°, since ν_1° induces λ_1°.

f_1 is also bounded, so that p and k, and therefore $p \circ h_1$ and $k \circ w_1$ are bounded. It follows that the integrals,

$$\int_{S_1} p \circ h_1 \, d\alpha_1 \text{ and } \int_{Q_1} k \circ w_1 \, d\beta_1,$$

are well-defined and finite. Hence $(p \circ h_1, k \circ w_1)$ is a measure-potential for ν_1° in the narrow sense. This implies that ν_1° is best for the modified allotment-assignment problem, ~~and~~ so ν° is best for the original. |||

This strengthens a previous result ~~(page)~~. Finally, we conclude with a theorem on the existence of optimal solutions.

Theorem: Let (S, Σ_s, α) and (Q, Σ_q, β) be measure spaces with $\alpha(S) = \beta(Q) < \infty$. Let Σ_s and Σ_q be the Borel fields of topologies T_s and T_q, respectively, these being topologically complete and separable. Let $h: S \to$ reals and $w: Q \to$ reals be semi-continuous, and let f: reals$^2 \to$ reals be bounded measurable with non-negative cross-differences.

Then ~~there~~ exists a best solution to the allotment-assignment problem of minimizing (91) over measures ν with marginals α and β.

Proof: By a preceding theorem (page), ~~there~~ exists a measure ν° with marginals α and β which satisfies measurable weight-falloff. By the theorem above, ν° is best. |||

(A) 8.6. Applications of Allotment-Assignment

The preceding section has been written at a rather abstract level, and the hurried reader may have trouble interpreting the results for the allotment-assignment problem in terms of a concrete Thünen system. We ~~shall~~ now give several illustrations, and show ~~in fact~~ that the allotment-assignment problem has applications well beyond the range of models contemplated in sections ~~§§1~~ 8.1 to ~~§§4~~ 8.4.

~~Let us~~ We begin with the original Thünen model. Here Space is a circular disc of finite radius r, its center being the nucleus. (S,Σ_s,α) is ordinary two-dimensional Lebesgue measure on the plane, restricted to S, so that the area of the entire system, $\alpha(S)$, is πr^2. The ideal distance of location s, $h(s)$, is simply the ordinary Euclidean distance of s from the nucleus. The set of land uses, Q, is finite — say $\{q_1,\ldots,q_n\}$. It is natural in this case to let Σ_q be the class of all subsets of Q. The ideal weight function w is now simply an n-tuple, $(w_1,\ldots,w_n)$, w_i being the weight of land use q_i, $i = 1,\ldots,n$. We ~~shall~~ assume to begin with that all ~~of~~ these w_i's are positive and distinct, and we may suppose that land uses are numbered in order of decreasing weight:

bold face Q

$$w_1 > w_2 > \ldots > w_n > 0.$$

The allotment measure β over (Q,Σ_s) is also characterized as an n-tuple $(\beta_1,\ldots,\beta_n)$, β_i being the allotment of the singleton set $\{q_i\}$. We assume that q28.2

$$\beta(Q) = \beta_1 + \ldots + \beta_n = \pi r^2 = \alpha(S). \qquad (8.6.1)\ \cancel{(1)}$$

That is, the acreage allotted to all activities together just uses up the area of the system. Equation (1) is a necessary condition for the existence of a feasible solution to the allotment-assignment problem with marginals α, β.

Now consider assignments ν on $(S \times Q, \Sigma_s \times \Sigma_q)$. Geometrically, $S \times Q$ consists of n replicas of the circular disc S, one for each q_i. Label these replicas $S_1,\ldots,S_n$. (Thus $S_i = S \times \{q_i\}$, which we may identify with S.) ν is then the direct sum of n measures $\nu_1,\ldots,\nu_n$, ν_i being over S_i. ν_i is, in fact, the areal distribution of land use q_i over Space. For feasibility one must have

$$\nu_i(S_i) = \beta_i, \qquad (8.6.2)\ \cancel{(2)}$$

$i = 1,\ldots,n$, and

$$\nu_1(F) + \ldots + \nu_n(F) = \alpha(F). \qquad (8.6.3)\ \cancel{(3)}$$

for all $F \in \Sigma_s$. Equation (2) states that the acreage occupied by each land use meets its allotment, while (3) states that the area of any region is just exhausted by its assignment to the various land uses. (The sets F on the left of (3) are replicas of the region $F \subseteq S$, and should, strictly speaking, be distinguished by subscripts i).

Now among these feasible ν's, consider one very special assignment: the Thünen assignment, ν^o. This is characterized by an n-tuple $(r_1,\ldots,r_n)$, where

$$\pi r_i^2 = \beta_1 + \ldots + \beta_i, \qquad (8.6.4)\ \cancel{(4)}$$

$i = 1,\ldots,n$. ν_i° is then defined to be ~~two~~ 2-dimensional Lebesgue measure truncated to the circular ring

$$\{s \mid r_{i-1} \le h(s) \le r_i\}, \qquad (8.6.5)\ \cancel{(5)}$$

$i = 1,\ldots,n$. (For $i = 1$, $r_0 = 0$). The feasibility of ν° is easily checked. ~~For example~~ e.g., the area of region (5) is $\pi r_i^2 - \pi r_{i-1}^2 = \beta_i$, which establishes (2).

It is customary to represent ν° by collapsing the replicas $S_1,\ldots,S_n$ onto the disc S. Measures $\nu_1^\circ,\ldots,\nu_n^\circ$ are mutually singular, and one can say, roughly, that land use q_i occupies the ring-shaped region (5). This gives the familiar picture of Thünen systems with concentric rings of land uses.

The Thünen assignment ν° satisfies the measurable weight-falloff condition. This is almost obvious, since weights fall as one moves outward from ring to ring. To prove it, let E_1, E_2 be two sets of positive ν°-measure, with E_1 southwest of E_2. This implies that

break at ↑

$$\nu^\circ(E_1 \cap S_i) > 0, \quad \nu^\circ(E_2 \cap S_j) > 0$$

for some i, j with $i > j$ (since $w_i < w_j$). Hence there must be points $(s_1, q_i) \in E_1$, $(s_2, q_j) \in E_2$, such that (s_2, q_j)

$$h(s_1) \ge r_{i-1} \ge r_j \ge h(s_2). \qquad (8.6.6)\ \cancel{(6)}$$

Relation (6) contradicts the assumption that E_1 is southwest of E_2, and the proof is complete.

#

18 Furthermore, ν^{o} is the only feasible assignment satisfying the measurable weight-falloff condition. This follows from the uniqueness theorem ~~(page)~~, whose premises are satisfied: Σ_q is the ~~sigma~~ σ-field inversely induced by w from the real Borel field, and

$$\alpha\{s \mid h(s) = x\} = 0 \qquad \text{(8.6.7)} \ \cancel{(7)}$$

for all real x. Equation (7) states that the area of the set of locations exactly at distance x from the nucleus is zero.

It follows (~~page~~) that the Thünen assignment is the unique feasible assignment ~~which~~ that minimizes total transportation costs, given by

$$\int_{S\times Q} h(s)w(q)\,\nu(ds,dq). \qquad \text{(8.6.8)} \ \cancel{(8)}$$

(Exercise: evaluate (8) explicitly in terms of $\beta_1,\ldots,\beta_n$, for $\nu = \nu^{o}$).

Furthermore, we can calculate the ~~potentials~~ potential, (p,k), explicitly for the Thünen assignment. Integral (8) is the special case of the allotment-assignment problem for which $f(x,y)$ is the product function xy. This is continuously differentiable, hence we may use the line-integral formulas, (5.83) ~~(83)~~ and (5.84) ~~(84) of section 5~~, to calculate p and k.

G28.3 The line of support, L, along which the integrals are to be taken, is a "staircase" polygon in the rectangle of support $X \times Y$, going horizontally from $(0,w_1)$ to (r_1,w_1), then down vertically to (r_1, w_2), then over to (r_2, w_2), etc., ending at (r_n, w_n).[21] Applying ~~(83) of section 5~~ (5.83), we find that p (whose domain is $X = \{x \mid 0 \leq x \leq r\}$) is a decreasing, convex, polygonal function, whose slope is $-w_i$ on the segment from r_{i-1} to r_i $(i = 1, \ldots, n)$; p is unique up to an additive constant.[22]

21

22

centered degree sign

The composition $p \circ h$ determines a real-valued function on Space~~, S,~~ which is a left half-potential for the allotment-assignment problem. This may be interpreted as land value — or, more precisely, as land-value density, taken per unit of ideal area ("dollars per acre"). Differences in land values at different sites reflect differences in locational advantages. In particular, the decline in land-value density at a rate of w_i per unit distance as one moves outward from the nucleus matches exactly the increase in transport cost incurred per unit ideal distance per unit ideal area.

Applying ~~(84) of section 5,~~ (5.84) we find that k (whose domain is $Y = \{y \mid w_n \leq y \leq w_1\}$) is an increasing, concave, polygonal function, whose slope is r_i on the segment from w_{i+1} to w_i, $(i = 1, \ldots, n-1)$. Actually, only the values of k at the points $w_1, \ldots, w_n$ are relevant, since the composite function $k \circ w$, which determines a right half-potential on Q, takes on only these values. The interpretation of k is also less clear than that of p, since k corresponds to the "artificial" allotment constraint, while p corresponds to the "natural"

n minus one

$3\frac{1}{2}\frac{1}{16}$ to 4

centered degree

areal capacity constraint. $k(w_i)$ may be thought of as the gross cost density associated with land use q_i, the sum of transport cost and land value per unit area. ¶ We have spelled out the application of the allotment-assignment analysis to this simple Thünen model in considerable detail. For the models ~~which~~ that follow ~~we shall be~~ the discussions are briefer, leaving the task of filling in the details of the argument to the reader.

Suppose the preceding Thünen system is modified by having a total allotment ~~which~~ that is less than the total areal capacity (change the middle equality sign of (1) to "<"). This is no longer in allotment-assignment form, but can be brought into this form by the device of adding an artificial "vacancy" land use, q_{n+1}, to the set Q. This q_{n+1} has an ideal weight, w_{n+1}, of zero, and is given an allotment, β_{n+1}, which just exhausts the surplus areal capacity.

The above analysis now applies verbatim to this modified system, and we find that q_{n+1} occupies the outermost ring ~~of the system~~, since it is the lightest land use. In other words, the occupied land is crowded in as closely as possible to the nucleus, the surplus vacant land being on the outskirts. The only other difference is that land-value density, $p \cdot h$, is constant in the vacant ring beyond radius r_n (since the slope of p is $-w_{n+1} = 0$ there).[23]

Concerning the meaning of "vacancy", it should be stressed that it refers to permanent vacancy; ~~— that is~~ i.e., over the entire time horizon of the system. Suppose, for example, that "Time" is a certain 100-year interval, and that a given site is unoccupied for 99 years, but ~~that it~~ does trade with the nucleus in the final year. The land use assigned to this site *is not* the vacancy land use, ~~The reason is that~~ because ideal weight involves an integration over all Time, and will therefore be positive in this case, whereas it must be zero for "vacancy".

Going a step further, one can allow some (or all) ~~of~~ the weights w_i to be *negative*. Formally, this creates no difficulties, and the unique minimizer of (8) is again the Thünen assignment: Concentric rings are occupied by land uses in order of decreasing weight, the outermost zone being occupied by the land use whose weight is most negative. The half-potential p is still convex but no longer decreasing; rather, it decreases at distances occupied by land uses of positive weight and increases at distances occupied by land uses of negative weight.

How does one interpret negative weights? One possibility is to think of these land uses as being *repelled* from the nucleus (say by pollution, crime, or other aspects of the urban syndrome) rather than as being *attracted* by transportation linkage. The land use of highest negative weight is the one most repelled, and it will naturally settle in the outermost ring.

(There is also a purely formal use for negative weights. Suppose the problem given is to *maximize*, rather than minimize, (8). This is equivalent to negating the weights and minimizing, which is again in allotment-assignment form. ~~In fact,~~ it follows that the solution to maximizing (8) is the anti-Thünen assignment, obtained by reversing the distance-ordering of land uses: the lightest one in the closest ring, the heaviest in the outermost ring, etc.)

We now drop the assumption that Space is a circular disc, and instead let it be any bounded measurable subset of the plane. Now α is still ~~two~~ 2-dimensional Lebesgue measure restricted ~~to~~ to S, ~~and~~ we still assume that $\beta(Q) = \alpha(S)$, and $h(s)$ is still the Euclidean distance from the nucleus to location s.

With these weaker assumptions, ~~it still turns out that~~ there is still a unique feasible assignment satisfying the measurable weight-falloff condition. This assignment has virtually the same form as the Thünen assignment ν^o; ~~Namely~~ viz., land use q_i will occupy part of a ring-shaped zone (centered on the nucleus) of the form (5), the q_i ranging outward in order of decreasing weight. The novelty is that not all the points of the plane satisfying (5) are available, but only those ~~which are~~ in S. (4) will also be false, in general. Instead, we have

$$\alpha[S \cap \{s | h(s) \leq r_i\}] = \beta_1 + \ldots + \beta_i \qquad \cancel{(9)}\ (8.6.9)$$

$i = 1,\ldots,n$. Here r_i, the distance of the borderline between the land-use zones for q_i and q_{i+1}, will generally be larger than the r_i of (4). One must go further out to get the same area, since pieces of the plane are missing.

The existence of a weight-falloff measure follows from the existence of $r_1, \ldots, r_n$ satisfying (9), and this in turn can be proved using the argument of (5.48)–(5.49) ~~(48)-(49) of section 5~~. ~~Also~~ The uniqueness theorem also still applies. This measure still minimizes (8). The potential (p, k) is the same as before. In short, removing an arbitrary piece of land from the system makes no qualitative difference in the solution.

The foregoing generalization applies to cases where a portion of the region is unavailable for occupancy. For example, this may be due to bodies of water, poor drainage, irregular terrain, or other natural adversities.[24] Or, some land may be ~~pre-emptied~~ preempted for public use or lie outside the legal jurisdiction of the system. For another application, suppose the system is subject to a zoning ordinance. This invalidates the allotment-assignment formulation, because some land uses can be assigned to some zones but not to others. However, if we confine attention to any <u>one</u> zone, ~~then~~ the formulation becomes revalidated, the set of land uses Q being replaced by Q', which is the subset of uses allowed in that particular zone. We may therefore expect that the Thünen ring pattern will be present <u>within</u> any particular zone, but that land uses, and the radii of the borderline between them, will vary from zone to zone.

~~Let us~~ We now drop the assumption that α is ~~two~~ 2-dimensional Lebesgue measure. Instead, we let it be <u>any</u> measure such that $\alpha(S) = \beta(Q)$ is finite, where S is still a bounded subset of the plane. (This could represent differential "fertility" of land, or the fact that land in some places has been augmented by the building of multiple-storfy structures). ~~It turns out that there~~ still exists a feasible assignment satisfying the measurable weight-falloff condition, and that it has virtually the same Thünen form as above. The only novelty that can arise would be if

$$\alpha\{s|h(s) = r_i\} > 0 \qquad (8.6.10)\ \sim\!\!\sim(10)\sim\!\!\sim$$

for some $i = 1,\ldots,n-1$, r_i being the radius of the borderline between land uses q_i and q_{i+1}. For suppose x acres of this borderline are to be assigned to activity q_i, where x is positive but less than (10). Then, provided α on $\{s|h(s) = r_i\}$ is not simply-concentrated, there will be more than one way to apportion the land of (10) between q_i and q_{i+1} without violating weight-falloff. Thus uniqueness *may* break down with this more general areal capacity measure α. Any of these weight-falloff measures is best for the allotment-assignment problem. The potential (p,k) is unchanged.

Q28.5

A slight modification of the procedure outlined above enables us to construct one of these Thünen assignments. ~~There~~ may not exist an r_i satisfying (9). Instead, we let r_i be the smallest number for which the right side of (9) does not exceed the left. If (10) obtains for ~~some~~ this r_i, assign to q_i that proportion of α on the borderline ~~which~~ that just fills out its allotment, and assign the rest to q_{i+1}. (If the allotment of q_{i+1} is also exhausted, assign the rest to q_{i+2}, etc.)

We ~~Let us~~ continue to generalize~~.~~ and ~~We~~ now abandon the assumption that $h(s)$ is the Euclidean distance from s to the nucleus. Instead, we let it be any bounded measurable function on S. $h(s)$, of course, is interpreted as ideal distance, and our present generalization amounts to breaking the identification between geographical and economic distance. Thus we can incorporate irregularities of terrain, ~~of~~ the transportation grid, tariffs not proportional to distance, zonal tariffs, etc.

Formally this generalization makes little difference, because the theory of the allotment-assignment problem makes no assumption that $h(s)$ has any relation to a metric on S. We cannot assume that there is a unique feasible weight-falloff measure, because the (10) phenomenon invalidates the uniqueness theorem. We can still assert the existence of such a measure, constructing it by the procedure outlined above. This measure still minimizes (8), and the potential (p,k) retains its properties.

But despite all this formal similarity, the geographical appearance of the resulting Thünen system can be radically altered. ~~The point is that~~ the land-use zones lie between contours of $h(s)$, as in (5). If $h(s)$ is Euclidean distance from the nucleus, these contours are concentric circles, and the familiar pattern emerges. With general $h(s)$, however, the zones can become quite irregular.

Let $\{s|h(s) = r_i\}$ be the "borderline" between zones occupied by land uses q_i and q_{i+1}. Then $q_1,\ldots,q_i$ occupy the open disc $\{s|h(s) < r_i\}$. (The borderline itself, if it has positive area, may be occupied by q_i, ~~or~~ by q_{i+1}, ~~or~~ maybe shared by them, etc.) The geographical shapes of the Thünen "rings" can then be gauged by examining the shapes of these open discs. This returns us to the discussion of section ~~3~~ 8. For the city-block metric — (generated perhaps by a rectangular road grid) — the open discs will be concentric diamonds, and the "rings" will be the regions between two diamonds at ideal distances r_{i-1} and r_i $(i = 1,\ldots,n)$. For the distance function determined by a system of high-speed radial arteries converging on the nucleus, the open discs will have an amoeboid shape, projecting outward along the arteries. We would then expect high-intensity land uses to "sprawl" along the arteries, while sites away from the arteries, at similar geographical distances from the nucleus, would have lower-intensity, more "rural" uses.

Limited-access transportation systems lead to ideal distances with disconnected open discs, and a corresponding fragmentation of the Thünen rings. The resulting geographical pattern is intriguing. Consider, for example, a plane ~~which~~ that is uniform except for a high-speed commuter railway connecting the nucleus with a series of isolated stops, s_1, s_2, s_3... in order of increasing distance. Away from the railway land uses will be arranged in concentric circular rings in order of decreasing land-use weight, q_1, q_2,... . Suppose the first stop, s_1, satisfies: $r_9 < h(s_1) < r_{10}$, so that points off the railway at that distance are in the zone of q_{10} (s_1 itself is geographically far beyond that zone, but the railway shrinks the economic distance). The weight-falloff condition then calls for a secondary sequence of concentric rings, centered on s_1, and beginning with land use q_{10} rather than q_1. The second stop, s_2, will be farther out, say at ideal distance between r_{16} and r_{17}. Then another secondary sequence will spring up around s_2, starting at q_{17}, q_{18},..., and so on for s_3, s_4,... .

The resulting pattern is a simple "urban hierarchy" organized in the form of a "metropolitan region". The "central" city ~~is the one that~~ grows up about the nucleus. "Satellite" cities grow up about the points of access to the transportation network. These cities lack, however, the full range of land uses; The heaviest, most intensive, most "urban" uses are missing. The further out the satellite is, the more uses will be missing. The incidence matrix, indicating which land uses are present in which cities, is of the "central place" type, in that, if a given land use occurs in a given city, all lower type land uses are also present (in this case, "lower" means "lighter").

q28.6 The One could make the transportation system itself hierarchical, which would lead to the satellite cities themselves spawning satellites, etc. The central place pattern would still obtain. The interesting point is that this rather complicated scheme follows from simple and plausible assumptions concerning the transportation system, coupled with the weight-falloff condition, which itself follows from the minimization of transport costs (8).

~~(8)~~

To round off the discussion of "non-Euclidean" ideal distances, it should be mentioned that the "borderlines" $\{s|h(s) = r_i\}$ can themselves be broad geographic zones. That is, changes in geographic distance lead to no change in transport cost over a certain range. This can occur in a thorough-going regime of freight absorption, or uniform charges over broad zones. This is quite common in postal and telephone rates, utility charges, retail deliveries, etc. Conversely, ideal distance can jump discontinuously, as at a tolling point or political border.

In all this, the pattern of land-value densities depends on ideal distance, and will therefore behave irregularly with respect to geographical distances. Thus land values will reach local peaks at points of access to the transportation system (highway interchanges, railway stops, etc.), will have ridges along radial arteries, ~~and so on~~ etc.

We have devoted almost all our attention so far to modifications in Space~~, S,~~ and its associates, α and h. Let us now turn to Q. This has been assumed finite with unequal weights, $w_1, \ldots, w_n$, with Σ_q being all subsets of Q.

~~Let us~~ now suppose that two different land uses have the same weight, say $w_1 = w_2$. Then uniqueness of a feasible weight-falloff assignment can, in general, no longer be guaranteed. Either q_1 or q_2 may occupy the innermost Thünen ring; more generally, any mixture of these over the innermost two zones ~~which~~ that satisfies their allotments will do. (Correspondingly, ~~of course,~~ some premise of the uniqueness theorem must fail. In this case it is the premise that Σ_q is the ~~sigma~~ σ-field inversely induced by w; ~~— since, for example,~~ e.g., the singleton $\{q_1\}$ is not the inverse image of any real Borel set.)

Let us now assume that Q is infinite. This encompasses the case, for example, where continuous variation in the intensity of a land use is possible. (Recall that intensity variations in "one" land use are to be formally considered as different land uses.) We have an allotment measure β on (Q,Σ_q) and a measurable weight function $w:Q \rightarrow$ reals, and we ~~shall~~ assume as usual that $\beta(Q) = \alpha(S) < \infty$. We ~~shall~~ also assume that different land uses have different weights. It is then natural to assume that Σ_q is the ~~sigma~~ σ-field inversely induced by w from the real Borel field. Also assume that $\{w(q)|q \in Q\}$ is a real Borel set. Finally, ~~— to make things simple —~~ assume that S, Σ_s, α, and h take the classical form, $h(s)$ being Euclidean distance to the nucleus (which is the center of circular disc S) and α being 2-~~two-~~dimensional Lebesgue measure.

Under these conditions, ~~there~~ exists exactly one feasible assignment $\nu°$ satisfying the measurable weight-falloff condition. Uniqueness follows at once from the uniqueness theorem. To prove existence (end of Section 8.5) ~~(page)~~, we construct topologies, T_s, T_q, on S and Q, respectively, making S and Q Borel subsets of separable and topologically complete spaces, with respective Borel fields Σ_s and Σ_q, and making h and w semi-continuous.

For T_s we choose the usual topology on the plane, restricted to S. For T_q we take all sets of the form $\{q|w(q) \in E\}$, E ranging over all *open* subsets of the real line with its usual topology. This collection is a topology over **Q**, the topology *inversely induced* by w. One verifies that all the stated conditions are satisfied.

This unique Thünen assignment is, intuitively, a limiting case of the classical pattern. We still have circular symmetry of land-use assignments, but do not necessarily have broad rings devoted to a single land use. Instead, there may be continuous variation of land uses as one moves outward — always going from heavier to lighter uses, of course. The potential (p,k) may still be constructed by the line-integral formulas ~~(83)~~ (5.83) and ~~(84)~~ (5.84) ~~of section 5.~~ Half-potential p is still convex (and decreasing, if w is positive); k is still concave and increasing. If p is differentiable at any given distance x, its slope is given by $-w$, where w is the weight of the land use located "at" distance x. The only novelty is that p and k need not have polygonal graphs, but may have continuously turning tangents over certain ranges.

g 28.4 As a final generalization, we can relax the *factorability* assumption (transport cost = ideal weight ~~times~~ times ideal distance), which underlies the objective function (8). Recall that the allotment-assignment problem, in its general form, calls for the minimization of

$$\int_{S\times Q} f(h(s),\ w(q))\nu(ds,dq), \qquad (11)$$

(8.6.11)

where f has positive (or non-negative) cross-differences. Integral (8) is merely the special case in which $f(x,y) = xy$. Now the measurable weight-falloff condition characterizes the optimal solutions to (11), in general. Hence all the qualitative features of the Thünen system carry over the more general situation of minimizing (11). The one novelty is that the potential (p,k) may lose its convexity or concavity.

This concludes our discussion of the application of the allotment-assignment problem to the classical Thünen model and its generalizations. The variety of situations covered is already considerable. We now generalize still further, departing more or less radically from the classical interpretation.

Allocation-of-Effort as a Thünen Problem

We have taken up in this book two broad types of optimization problems: the allocation-of-effort problem of chapter 5 and the transportation problem of chapter 7 (of which the allotment-assignment problem is a special case). These problems are quite different in construction; but there is one kind of situation to which they both apply, each in its own way.

Consider the following search problem. One is prospecting an unexplored region for minerals, and ~~one is to~~ distributing searching effort over Space so as to maximize expected return.

~~There is~~ a single location which is the "base of operations" — say, the railhead ~~which~~ that connects one with civilization. Now, even if the prospect of finding minerals is uniform over the region, it seems reasonable that ~~one~~ a person should search more intensively in the vicinity of the base than far away from it. For greater transport costs are incurred at distant points; The process of exploration itself is more costly, and so is the shipment of any minerals ~~which~~ that are discovered.

~~Now~~ this search problem is of the allocation-of-effort type of chapter 5. Yet the optimal solution appears to be a pattern ~~which~~ that is symmetric about the base of operations, such that intensity falls off with increasing distance from the base. This certainly mimics the Thünen pattern. The question then arises: Can we associate an allotment-assignment problem with the original allocation-of-effort problem, such that their optimal solutions correspond in some way? This would provide insight into the phenomenon of the Thünen pattern arising in an apparently very different type of problem.

Before going into this, ~~let us~~ we give a few more examples. Consider a farmer faced with the problem of distributing fertilizer over his fields. The farmhouse provides the "base of operations". If there are no geographic irregularities, the distribution ~~which~~ that maximizes total return would again appear to conform to the Thünen pattern, with more intensive fertilizer use on the nearer fields.

Again, consider a discriminating monopolist, with a single factory located at s_o. ~~We~~ suppose that consumers are uniformly distributed over Space, and ~~that~~ all have identical demand functions for the commodity produced at s_o. Assume that the monopolist pays for transportation, and that he can freely vary delivered price from one location to another – or, what is the same thing, that he can freely choose the density pattern of deliveries over Space, $\delta: S \to$ reals. (δ is measured in units of, say, tons per year per acre). The profit-maximizing density would then appear to have the Thünen pattern, in the sense that δ should decrease with increasing distance from s_o.

As a simple generalization of this case, drop the assumption that consumers are uniformly distributed. Then we still reach the same conclusion, provided $\delta(s)$ is interpreted as the density of deliveries *per person*, rather than *per acre*. A non-rigorous argument for the "intensity-falloff" of δ goes as follows. Suppose there were sites, s_1, s_2, with s_2 further from the factory than s_1, but with $\delta(s_2) > \delta(s_1)$. Then interchange these densities between one person at s_1 and one person at s_2. This leaves gross revenue intact, but reduces transport costs; hence the original distribution was non-optimal.

As a last example, let some public-service facility – such as a police or postal station – be located at s_o. This is to serve a certain hinterland and the question is: How shall the services be spatially distributed over this region? (Intensity of service could be measured, say, by frequency of police patrols or mail pickups.) With uniformly distributed population, and with uniform benefits as a function of service intensity, it again seems reasonable that intensity of service should decline with distance from s_o, to maximize total benefits net of transportation costs.

We now proceed to the analysis. The allocation-of-effort problem is determined by a ~~sigma~~ σ-finite measure space, (S, Σ_S, α), three measurable functions, $u: S \times \text{reals} \rightarrow \text{reals}$, and $b, c: S \rightarrow \text{extended reals}$, and two extended real numbers, $L_\circ$ and L°. The problem is to choose a measurable function $\delta: S \rightarrow \text{reals}$ to maximize

$$\int u(s, \delta(s))\, \alpha(ds) \qquad \text{(8.6.12)}\ \cancel{(12)}$$

subject to the constraints

$$L_\circ \leq \int_S \delta\, d\alpha \leq L^\circ \qquad \text{(8.6.13)}\ \cancel{(13)}$$

~~and~~

$$b \leq \delta \leq c. \qquad \text{(8.6.14)}\ \cancel{(14)}$$

Here (12) is an indefinite integral over S, and "maximization" is to be understood in the sense of standard ordering of pseudomeasures. (If (12) is well-defined and finite as a definite integral for all feasible δ, this ~~of course~~ reduces to the ordinary size comparison of definite integrals.) It is also required that $\int_S \delta\, d\alpha$ be finite, even if $L_\circ$ or L° is infinite. (For convenience we have changed notation somewhat from that of chapter 5.)

We now identify a special subclass of the problems (12)–(14), one ~~which~~ that embraces all the examples cited, and for which the Thünen pattern emerges. The first special assumption is that $\underline{b}$ and $\underline{c}$ be constants (one or both may be infinite). The second special assumption is that

$$\underline{u}(\underline{s},\underline{x}) = -\underline{f}(\underline{h}(\underline{s}),\ \underline{x}), \qquad (8.6.15)\ (15)$$

for all $\underline{s} \in \underline{S}$, $\underline{x}$ real, for some measurable functions $\underline{h}:\underline{S} \to$ reals, $\underline{f}:\text{reals}^2 \to$ reals, where $\underline{f}$ has positive cross-differences. There are a few more minor technical assumptions, but these are the two major ones. ~~Let us~~ We now interpret them.

reals^2

The first states that the limits on the range of permissible densities do not vary from point to point. This is plausible for all ~~of~~ our examples. In fact it is reasonable to take $\underline{b} = 0$, since negative densities are meaningless for them, and ~~there~~ is no positive minimal density required. (Exception: in the case of public services, there may be an institutional requirement to reach some social minimum for all regions or persons. In this case, $\underline{b} > 0$, but still constant). As for the upper limit $\underline{c}$, there may be either no such limit ($\underline{c} = \infty$), or there may be some uniform "saturation level" of intensity, for regions or persons.

As for the second assumption, the special form (15) is clearly reminiscent of the integrand in the allotment-assignment problem, (11). (The minus sign is inserted because the objective in (12) is to maximize, whereas one minimizes in allotment-assignment). ~~Let us~~ We show that (15) is a plausible form for the objective in each of our examples. In evaluating a density pattern there are two factors to consider: the *gross returns* over Space ~~which~~ accruing from that pattern, and the *transportation cost* ~~which is~~ incurred by that pattern. In the exploration example, gross return would be the expected payoff from mineral~~s~~ discoveries (less searching costs other than transportation). For the farmer, gross return is ~~his~~ profit on crops grown, less the cost of fertilizer. For the monopolist, it is ~~his~~ revenue from sales. And, for the public service facility, it is the social benefit from the service rendered.

g 28-8

~~Now~~ in all these cases the gross return at a point will depend only on the intensity at that point, not on the location of the point *per se*. This results from our uniformity assumptions. The gross return is represented by a measurable function g: reals → reals. As for transportation cost, ~~let us, to begin with,~~ we make the simplest factorability assumption: transport cost incurred at point s is the product of intensity at s and the distance, $h(s)$, of s from the base of operations. We then have

$$u(s,x) = g(x) - h(s)\cdot x \qquad (8.6.16) \quad \cancel{(16)}$$

The critical observation to make is that (16) is of the form (15) (with f having positive cross-differences) regardless of what the function g is. This follows at once from the fact that g does not depend on s. We may also discard the special transport cost assumption, and assume more generally that transport cost is of the form $f_1(h(s), x)$, for some measurable function f_1 with positive cross-differences. Throwing in $g(x)$ does not alter this property, and (15) is still valid. Thus all the examples given ~~do~~ seem to be encompassed by our special assumptions.

We now want to translate from the language of the allocation-of-effort problem to the language of the allotment-assignment problem. The latter speaks of "land uses". We now let each possible intensity level be a land use. The set of land uses, Q, may be formally identified with the real line. (Actually, the interval $[b,c]$ would suffice). This is a radically stripped-down version of the full-blown land-use concept. For the latter, the answer to the question, "what's going on here?", would require at least the specification of two measures over $R \times T$: production and consumption. ~~For us~~ here, it requires just the specification of a real number, indicating intensity (intensity of search, ~~or of~~ sales, ~~or~~ fertilizer use, etc.). The weight-function w is simply taken to be the identity, and may be ignored.

Choice in the allotment-assignment problem is over measures ν on $(S \times Q, \Sigma_s \times \Sigma_q)$; in the allocation problem, choice is over measurable functions $\delta: S \to$ reals. To translate from the latter to the former, use the formula

$$\nu(G) = \alpha\{s \mid (s,\delta(s)) \in G\}, \qquad (8.6.17) \ \cancel{(17)}$$

for all $G \in \Sigma_s \times \Sigma_q$. ($\Sigma_q$ is the Borel field on the real line Q). Here α is the measure of (12)–(13), and has the interpretation of acreage, ~~or~~ population, etc. That is, given an intensity distribution represented by δ, (17) shows how to express it as an assignment measure ν. For example, let G be the rectangle $E \times F$, and let α be areal measure. Then $\nu(E \times F)$ is the area in region E ~~which is~~ assigned to intensities among the numbers F, and this is just what the right side of (17) equals.

To see that ν defined by (17) is indeed a measure, note that $s \to (s,\delta(s))$ is a measur~~ab~~able mapping from S to $S \times Q$. Equation (17) states that ν is the measure on $S \times Q$ induced by this mapping from α on S. (There is one fine point here: The solutions to the allocation problem are not really the densities δ, but the indefinite integrals $\int \delta \, d\alpha$ they represent. Two densities, δ_1 and δ_2, represent the same indefinite integral iff they are identical α-almost-everywhere. Equation (17) would hardly be satisfactory if two equivalent δ's yielded different ν's. But ~~in fact~~ they yield the same ν, as one may verify.)

In the allotment-assignment problem, ν is feasible iff it has the prescribed marginals, α and β. As our notation indicates, we are taking over the measure space (S,Σ_s,α) of the allocation-of-effort problem bodily to be the left marginal space of the allotment-assignment problem. We now show that any assignment ν defined by (17) automatically satisfies the left-marginal allotment-assignment constraint. In fact, for any $E \in \Sigma_s$,

break at

$$\nu'(E) = \nu(E \times Q) = \alpha(E),$$

from (17), so that α is the left marginal of ν. We have not yet defined the allotment measure β on (Q, Σ_q); this will be ~~taken up~~ later.

428-9

#

Theorem: Let (S, Σ_s, α) be a non-atomic ~~sigma~~ σ-finite measure space; let $h{:}S \to$ reals and $f{:}\text{reals}^2 \to$ reals be measurable, with f having positive cross-differences, and $f(y,\cdot)$ being upper semi-continuous for each real y; let b, c, L_o, and L^o be four extended real numbers.

(raised o)

D

Let δ^o be unsurpassed for the problem of minimizing

$$\int f(h(s), \delta(s))\,\alpha(ds) \qquad (8.6.18)\ \cancel{(18)}$$

over the set of measurable functions $\delta{:}S \to$ reals satisfying

$$L_o \le \int_S \delta\, d\alpha \le L^o \qquad (8.6.19)\ \cancel{(19)}$$

(the integral in (19) being well-defined and finite), and

$$b \le \delta \le c. \qquad (8.6.20)\ \cancel{(20)}$$

Then ν^o, defined by (17) for $\delta = \delta^o$, satisfies the (measurable) weight-falloff condition (with respect to h and w, the latter being the identity function).

#

Proof: The premises imply that ~~there exists~~ an extended real number, p_o, and a set $E_o \in \Sigma_s$ exist, such that $\alpha(E_o) = 0$, and, for each $s \in S\backslash E_o$, $\delta^{\underline{o}}(s)$ minimizes

$$f(h(s),\ x) + p_o x \qquad (8.6.21)\ \cancel{(21)}$$

$p_o X$

over the real numbers x in the closed interval $[b,c]$ (page). Suppose first that p_o is finite, and let $s_1,\ s_2 \in S\backslash E_o$. Using the abbreviation f_{ij} for $f(h(s_i),\ \delta^{\circ}(s_j))$, $i = 1,2$, $j = 1,2$, we then have

$$f_{11} + p_o\delta^{\circ}(s_1) \leq f_{12} + p_o\delta^{\circ}(s_2)$$

~~and~~

$$f_{22} + p_o\delta^{\circ}(s_2) \leq f_{21} + p_o\delta^{\circ}(s_1).$$

These yield

$$f_{11} + f_{22} \leq f_{12} + f_{21}. \qquad (8.6.22)\ \cancel{(22)}$$

~~Now~~ It is impossible that both $h(s_1) < h(s_2)$ and $\delta^{\underline{o}}(s_1) < \delta^{\underline{o}}(s_2)$ be true, for in this case (22) would violate the positive cross-differences condition. Hence

$$h(s_1) < h(s_2) \quad \text{implies} \quad \delta^{\circ}(s_1) \geq \delta^{\circ}(s_2). \qquad (8.6.23)\ \cancel{(23)}$$

Now suppose ~~that~~ $p_o = +\infty$. Minimization of (21) then means (by convention) that x is to be as small as possible, so that $\delta^{\underline{o}}(s) = b$ for all $s \in S\backslash E_o$. (23) then holds automatically. A similar conclusion follows for $p_o = -\infty$. We conclude that (23) is true for any $s_1,\ s_2 \in S\backslash E_o$.

~~Let us~~ We now go to the product space, $(S \times Q, \Sigma_s \times \Sigma_q)$. Let G_1, G_2 be two measurable sets, with G_1 southwest of G_2; ~~That is,~~ i.e. if $(s_i, q_i) \in G_i$, $i = 1,2$, then $h(s_1) < h(s_2)$, and $q_1 < q_2$. (Remember that Q is the real line). Define two new sets, H_1, H_2, by

$$H_i = G_i \setminus (E_o \times Q),$$

$i = 1,2$. At least one of the sets H_1, H_2 must own no point of the form $(s, \delta^o(s))$, for otherwise (23) would be violated. Hence, from (17),

$$\nu^o(H_i) = \alpha(\emptyset) = 0$$

must be true for at least one index $i = 1,2$. We also have

$$\nu^o(E_o \times Q) = \alpha(E_o) = 0$$

so that

$$\nu^o(G_i) \leq \nu^o(H_i) + \nu^o(E_o \times Q) = 0$$

must be true for at least one index $i = 1,2$. Thus ν^o satisfies measurable weight-falloff. $\square$

This result ratifies our intuition concerning the nature of optimal solutions to the allocation problems discussed. They do indeed yield a Thünen system when translated by the natural formula (17) into the proper language. The assumptions made, in addition to the two already discussed, are that α be non-atomic and that $f(y,\cdot)$, as a function of its second argument, be upper semi-continuous for each real y.

9a1 39-1 ctr dot

10/12 x 27 TR

We mention a few generalizations. As discussed in ~~chapter~~ Section 5.4, the non-atomicity assumption on α could be dropped, at the cost of assuming $f(h(s), \cdot)$ to be convex for all s in the atomic part of S. (If α is population measure, the atomic part may be thought of as the "cities"; if α is areal measure, it may safely be assumed to be non-atomic). For the objective function of type (16), convexity of f is the same as concavity of g (= diminishing marginal returns to intensity of effort), which is not implausible but somewhat restrictive. Upper semi-continuity of f is the same as lower semi-continuity of g, which is so weak as to amount to no assumption at all from the practical point of view.

The assumption that b and c are constants can also be weakened to the following: For any $s_1, s_2 \in S$, if $h(s_1) < h(s_2)$, then $b(s_1) \geq b(s_2)$ and $c(s_1) \geq c(s_2)$. The preceding proof still goes through, with some minor complications whose discussion we omit.

It remains for us to complete the construction of the allotment-assignment problem to be derived from the allocation-of-effort problem (18)–(20). We make the additional assumption at this point that α is finite, and define the allotment measure β on the real line (Q, Σ_q) by

$$\beta(F) = \alpha\{s \mid \delta^{o}(s) \in F\}, \qquad \text{(8.6.24)} \; \cancel{(24)}$$

for all Borel sets F. This is the measure induced on the real line from α by the function δ^{o}. (If α were infinite, one could not guarantee the ~~sigma~~ σ-finiteness of β; in particular, β would not be ~~sigma~~ σ-finite if δ° were constant).

#

Theorem: Assume all the premises of the preceding theorem, and in addition assume that f and α are bounded. Consider the following allotment-assignment problem on the space $(S \times Q, \Sigma_s \times \Sigma_q)$: Minimize

$$\int_{S\times Q} f(h(s),\, q)\, \nu(ds, dq) \tag{25}$$

(8.6.25)

over the set of measures ν on $S \times Q$ ~~which~~ that have left and right marginals α and β, respectively. (β is given by (24), where δ° is unsurpassed for the problem (18)–(20).)

Then the assignment ν°, defined by (17) for $\delta = \delta^{\circ}$, is best for this problem.

#

Proof: First, ~~one~~ we immediately ~~verifies~~ verify that ν° has α and β for its left and right marginals, respectively, so that it is feasible. By the preceding theorem ν° satisfies measurable weight-falloff, and this fact, together with the boundedness of f and α (hence β), implies that ν° is best (page 000). |||

#

This artificially constructed allotment-assignment problem has the following intuitive meaning. Start with the density pattern δ° which is unsurpassed for the allocation-of-effort problem. Now consider any "reshuffling" of the pattern ~~which~~ that preserves the distribution of density aggregated over all ~~of~~ S. ~~(For example,~~ e.g., if $\alpha\{s\,|\,\delta^{\circ}(s) < 20\} = 100$, ~~then~~ the reshuffled pattern will also have 100 acres with density under 20, though ~~of course~~ the actual region of low density may be different). These reshufflings remain feasible for the allocation-of-effort problem, hence none ~~of them~~ can surpass δ^{o}. Now in the associated allotment-assignment problem, the allotment measure β is precisely this density distribution, and the corresponding constraint assures that, in a sense, the feasible assignments ν are "reshufflings" of ν^{o}.

The boundedness of f and α insures that (18) is well-defined and finite as a definite integral. In fact, it is equal to (25) if ν is derived from δ by (17). Hence the objective functions of the two problems coincide. The allotment-assignment problem is in a sense a subproblem of the original allocation-of-effort problem, in that it imposes the extra allotment constraint That ~~which~~ restricts comparison to reshufflings. ~~As previously discussed,~~ this is the general rôle of the "artificial" allotment constraint.[25]

Thünen Systems Without a Nucleus

Until now ~~Up to this point,~~ the nucleus, or "base of operations", has played a crucial role in the interpretation of the allotment-assignment problem, since the ideal distance $h(s)$ has been taken to be $h(s, s_N)$, the distance between location s and the nucleus s_N. We now show that a considerably more general interpretation is possible, one that ~~which~~ need not single out any particular location for special treatment.

9-2 The idea is this. We have assumed that any point $\underline{s}$ must trade exclusively with nucleus s_N. Suppose instead that the pattern of trade of a point is given by a distribution, ρ, over Space. $\rho(S) = 1$ — (i.e. ~~that is~~, ρ is formally a probability measure) — and $\rho(F)$ is taken to be the fraction of total trade (exports plus imports, measured by ideal weight) of s which terminates in region F, for all measurable F. The "nuclear" case is precisely that in which ρ degenerates to a measure simply-concentrated at s_N.

For example, there may be several nuclei — (say, $s_{N1}, \ldots, s_{N_k}$) and trade is to be divided among them in the given proportions $\rho_1, \ldots, \rho_k$. Or, ρ may be non-atomic, proportional perhaps to the distribution of population over Space. In any case, ρ is given as a condition of the problem.

We ~~Let us~~ now generalize still further, by allowing the spatial distribution of trade to depend on location. To represent this, we take ρ to be a conditional probability measure, with domain $S \times \Sigma_s$. i.e. ~~That is~~, for any $s \in S$, $\rho(s,\cdot)$ is a measure over S, with $\rho(s,S) = 1$, and, for any $F \in \Sigma_s$, $\rho(\cdot,F)$ is a measurable function. The interpretation is: $\rho(s,F)$ is the fraction of total trade of location s that ~~which~~ terminates in region F.

This allows a good deal of flexibility and realism to be incorporated into the conditions of the problem. We expect in a general way, for example, that locations tend to trade heavily with regions near themselves. One restriction is essential, however, for the analysis ~~which~~ that follows. Though distribution may depend on location, it must not depend on the land use whose exports and imports are being distributed.

It remains to show the conditions under which this more general scheme still leads to an allotment-assignment problem. Feasibility conditions are the same: There is an areal measure α on (S, Σ_s), and an allotment measure β on (Q, Σ_q) which are the marginals of any feasible assignment ν on $(S \times Q, \Sigma_s \times \Sigma_q)$. ~~As~~ As for the objective, we still postulate a weight-function $w: Q \to$ reals. But the distance function $h: S \to$ reals seems to be missing, since there is no nucleus. Instead, we go back a step, and assume that unit transport cost between any two locations can be expressed by a measurable function $g: S_1 \times S_2 \to$ reals. (S_1 and S_2 are both identical to S; the subscripts are inserted for clarity). Function g need not obey any of the postulates for a metric, except for symmetry: $g(s_1, s_2) = g(s_2, s_1)$.

The total transport cost incurred by an assignment ν is then

$$\int_{S_1 \times Q} \nu(ds_1, dq) \int_{S_2} \rho(s_1, ds_2)\, w(q) g(s_1, s_2). \qquad \text{(8.6.26)} \ \cancel{(26)}$$

Integrating from right to left in (26), the integration over S_2 yields the transport cost (per ideal acre) incurred in the process of distributing the exports and imports of land use q at s_1, according to the spatial pattern $\rho(s_1,\cdot)$. But if we now define the function $h: S \to$ reals by

$$h(s) = \int_{S_2} g(s,s_2)\,\rho(s,ds_2), \tag{27}$$

then h is measurable, and (26) is equal to

$$\int_{S\times Q} h(s)w(q)\,\nu(ds,dq), \tag{28}$$

which is in allotment-assignment form.[26]

Thus our generalized interpretation has the same formal structure as before, provided we define "ideal distance" by the special rule (27). Let us check to see what happens when there *is* a nucleus, s_N. In this case we have $\rho(s,F) = 1$ if $s_N \in F$, and $= 0$ if $s_N \in S\backslash F$, all $s \in S$, $F \in \Sigma_s$. The integral (27) then reduces to $g(s,s_N)$, and this is indeed our definition of $h(s)$ in the "nuclear" case.

The optimal solutions to the allotment-assignment problem with the objective of minimizing (28) must of course satisfy the measurable weight-falloff condition with respect to (h,w). Thus the heavy land uses will be assigned to regions of low h-values. In the "nuclear" case this has the geometric interpretation that the heavy land uses crowd in about the nucleus. What interpretation offers itself in the general, non-nuclear, case?

gal 29-3

We ~~shall~~ discuss this under the assumption that ρ is independent of s. ~~That is,~~ i.e. ~~there is~~ a fixed measure ρ over Space determin~~ing~~es the distribution of exports and imports from any location with any land use. Also assume that g is a metric on S. ~~Now~~ in the nuclear case, the nucleus can be characterized as the site for which h is minimal ($h(s_N) = 0$, and h is otherwise positive). This suggests looking for a site s_o ~~which~~ that minimizes $h(s)$ of (27). Indeed, the weight-falloff condition requires that heavy land uses crowd in about s_o just as they do about the nucleus when the latter exists.

~~Now~~ the problem of finding a location s_o that minimizes (27) is ~~a~~ basic ~~one~~ in spatial economics. This is the Weber problem – (to be discussed in detail in chapter 9) – and an optimal location is a Weber point or median of the distribution ρ. (Remember that ρ is independent of s; hence it may be thought of as simply a measure, not a conditional measure). Thus land uses will tend to arrange themselves in a pattern ~~which~~ that mimics the nuclear case, the median of ρ playing the rôle of "pseudo-nucleus". The difference is that the median may be otherwise just like any other point, ~~which~~ with no tendency for transportation flows to concentrate on it. Also, the points on a borderline between successive land uses will not generally be equidistant from the median, as they would be from the nucleus.[27]

27

B

Quality Complementarity

It is a matter of everyday observation that rich people tend to live in better housing than poor people, that good students tend to go to good schools, that abler managers hold more responsible positions, and that "desirable" husbands tend to marry "desirable" wives. Do these cases have anything in common?

~~In the~~ first ~~place,~~ they refer to associations of two kinds of entities: people and housing, students and schools, workers and jobs, husbands and wives. ~~In the~~ second ~~place~~, each of these two kinds of entities are ordered on a "quality" scale of some sort – (~~whether~~ by wealth, ~~by~~ ability, ~~by~~ aesthetic appeal, etc.). And third~~ly~~, out of all the possible ways that entities of various qualities could associate with each other, those of "high" quality on one scale tend to associate with those of "high" quality on the other, and similarly the "lows" associate with the "lows".

These characteristics establish a formal link with Thünen systems. Here the two kinds of entities are locations and land uses – i.e. ~~that is~~, the points of S and Q, respectively. The scaling of these entities is accomplished by the distance and weight functions, h and w, respectively. (For vividness, think of heavy land uses as being of "high" quality, and similarly for "close", low-h, locations. ~~Note that~~ site quality varies inversely with h; this inversion is needed to conform to the above usage). ~~And~~ the high-high, low-low quality association is represented precisely by the weight-falloff condition ~~which~~ that characterizes optimal land-use assignments ν.

We ~~Let us~~ illustrate this last point by the housing example. Interpret S as the set of housing types, and Q as the set of family types. An assignment of families to housing is represented by a measure ν over $S \times Q$; i.e. $\nu(E \times F)$ is ~~, say,~~ the number of square-feet of housing of types E occupied by families of types F. ~~Now~~ let $w(q)$ be the wealth of family type q, and let $h(s)$ be an index of quality of housing type, low h corresponding to high quality ("h" stands for "humbleness"). Then the fact that wealthy families occupy high-quality housing is expressed by the measurable weight-falloff condition on ν: if G_1 is southwest of G_2 (so that $(s_i,q_i) \in G_i$, $i = 1,2$, implies s_1 is of higher quality and q_1 of lower wealth than s_2, q_2, respectively), then either $\nu(G_1) = 0$ or $\nu(G_2) = 0$.

This all suggests that the analytical apparatus we have developed in this chapter may serve as an explanation of these rather diverse phenomena. ~~Now~~ this apparatus so far has run exclusively in terms of optimization (specifically, the minimization of cost in the allotment-assignment problem). Some of the situations ~~we have~~ mentioned may be interpreted directly as optimization problems. For example, a firm has an executive staff of varying ability, and a set of positions of varying responsibility, and ~~it is~~ must to fill the latter with the former ~~so as~~ to maximize profits. But more often we are talking about social equilibrium situations, where no single ~~will~~ person decides the final outcome. Even here, it is often enlightening to express the equilibrium as the solution to an optimization problem, though no one is consciously trying to solve this problem. (In ~~the following~~ section 8.7, we ~~shall~~ connect the "classical" Thünen equilibrium with the allotment-assignment problem in this way).

~~Let us~~ We examine both ~~of~~ these approaches for the situations ~~we have~~ mentioned. For the optimization approach, it is convenient to choose a slightly altered form of the allotment-assignment problem, in which one maximizes rather than minimizes. Specifically, one is to maximize

$$\int g(-h(s),\ w(q))\,\nu(ds,dq) \qquad (8.6.29)\ \cancel{(29)}$$

over feasible assignments ν. Here $w(q)$ is the quality index of $q \in Q$, and $-h(s)$ is the quality index of $s \in S$. (The minus sign is used to facilitate comparison with the original form of the problem).

If we define the function f:reals$^2 \to$ reals by

$$f(x,y) = -g(-x,y), \qquad (8.6.30)\ \cancel{(30)}$$

~~one easily~~ we can see~~s~~ that the preference ordering determined by (29) is the same as that determined by minimizing

$$\int f(h(s),\ w(q))\,\nu(ds,dq), \qquad (8.6.31)\ \cancel{(31)}$$

which is the original allotment-assignment objective function.

Now the critical feature of the allotment-assignment problem is that f in (31) should have positive cross-differences. By (30), this is true iff the function g has positive cross-differences.

What is the concrete significance of this property? One may speak of positive cross-differences of g in (29) as expressing complementarity between the two quality indexes, w and -h. ~~Indeed,~~ An older generation of economists were in the habit of defining complementarity between two commodities by the sign of their cross-derivative in someone's utility function: X and Y are complementary iff

$$D_2[D_1 g(x,y)] \geq 0 \qquad (8.6.32)\ \cancel{(32)}$$

for all x,y. Now one may show, provided $D_2[D_1 g]$ exists and is finite, that g has non-negative cross-differences iff (32) holds.[28] (When utility came to be interpreted as merely an indicator of preference, the definition (32) had to be abandoned. For the same preference ordering might have two utility indicators, one satisfying (32) and the other not.[29] This objection does not apply to g of (29).)

Consider the job-assignment example. A firm has a managerial staff, graded by ability, and positions to be filled, graded by responsibility. Suppose ~~that~~ g(x,y) is the profit generated by a ~~man~~ person of ability $\underline{x}$ placed in a position of responsibility $\underline{y}$, total profit generated by an assignment ν being given by (29). ~~Now~~ it is plausible that ability differentials show up more strongly, the more responsible ~~is~~ the position to be filled. Thus brilliant $\underline{A}$ might outshine mediocre $\underline{B}$ in a leadership position, but not do much better as a subaltern, because the latter position does not afford much scope for $\underline{A}$'s talents. Letting $x_1 < x_2$ be two ability levels, and $y_1 < y_2$ two job responsibility levels, the assumption just stated is that

$$g(x_2,y_2) - g(x_1,y_2) > g(x_2,y_1) - g(x_1,y_1). \qquad (33)$$

(8.6.33)

But this is precisely the positive cross-differences condition on $\underline{g}$!

Again, to take the school-assignment example, let $g(x,y)$, measured in dollars, be the "social benefit" from having a student of ability level $\underline{x}$ attend a school of quality level $\underline{y}$. Condition (33) then states that the differential benefit in favor of the ~~more~~ abler student is greater at the higher quality school. This again is ~~a~~ not an implausible assumption. If we now imagine a coordinated school policy aimed at assigning students ~~so as~~ to maximize overall social benefit, it will have an allotment-assignment problem to solve. The optimal solution will then be, under rather general conditions, a weight-falloff measure, which means that the better students go to better schools.

To illustrate how a "weight-falloff" assignment might be a social equilibrium, consider the housing example. Again we make the unrealistic assumption that there is a single "quality" dimension along which housing types can be arrayed. Let s_1 and s_2 be two housing types, s_1 having the higher quality: (say s_1 has a scenic view ~~which~~ that s_2 lacks, or s_1 has central air conditioning, etc.). A given family is willing to pay a premium to occupy housing type s_1 rather than s_2. We now assume that, the wealthier the family, the greater the premium it is willing to pay for the quality differential. This highly plausible assumption leads to the "weight-falloff" equilibrium. The basic argument can be illustrated in the case where there are just two wealth levels: "rich" and "poor". Suppose "rich" families are willing to pay \$100 to occupy s_1 with its scenic view rather than s_2, while "poor" families are willing to pay just \$10 for this privilege. It cannot then happen that in equilibrium there are both rich families living in the lower quality housing s_2 and poor families living in the higher quality housing s_1. For if the rent differential exceeds \$10, the poor families in s_1 do better to switch to s_2; while if the rent differential is less than \$100, rich families in s_2 do better to switch to s_1. Since one of these cases must occur, somebody is out of equilibrium. We conclude that the rich occupy high- quality housing and the poor occupy low-quality housing in equilibrium.

~~Furthermore,~~ this situation is associated in a natural way with the following "artificial" allotment-assignment problem. Let $x_1 > x_2$, and choose a function g: reals$^2 \to$ reals such that

$$g(x_1, y) - g(x_2, y) \qquad (8.6.34)\ \cancel{(34)}$$

equals the premium ~~which~~ families of wealth y are willing to pay to occupy housing of quality index x_1 rather than x_2. The assumption we made above is that the difference (34) increases with y for fixed x_1, x_2; this is the same as saying that g has positive cross-differences.

~~Now~~ let measure α on the set of housing types S, and measure β on the set of family types Q be given by: $\alpha(E)$ = square-feet of housing of types E and $\beta(F)$ = square feet of housing occupied by family types F in the above social equilibrium ν°. Then ν° will ~~then~~ be the optimal solution to the problem of maximizing (29) — the integrand g satisfying (34) — over assignments with marginals α and β. Furthermore, the prices associated with the various housing qualities turn out to be a left half-potential for this problem.

Note that (34) does not determine g uniquely. Indeed, adding to g an arbitrary function ~~which~~ that depends only on y will not affect (34). But this transformation does not alter the preference order determined by (29), hence it yields essentially the same allotment-assignment problem. One should be cautious in drawing normative conclusions from the fact that ν°

optimizes the problem just constructed: The market implicitly weights the preferences of different families, and this weighting need not coincide with that derived from some ethical principle.

We have run through the foregoing analysis rather rapidly, because ~~the next~~ section, 8.7, will cover the same ground more elaborately in the context of Thünen systems proper.

(W)

(B) A Combinatorial Application

Given 2n real numbers, $x_1 < x_2 < \ldots < x_n$, and $y_1 < \ldots < y_n$, consider the problem of minimizing the sum

$$x_1 y_{\pi(1)} + x_2 y_{\pi(2)} + \ldots + x_n y_{\pi(n)} \tag{8.6.35}$$

over all permutations π of $\{1,\ldots,n\}$. According to a theorem of Chebishev, the unique minimum occurs when the y's are taken in reverse order, matching y_n with x_1, etc. This is easily proved. For any other permutation π, there is an index j such that $y_{\pi(j)} < y_{\pi(j+1)}$. But then

$$x_j y_{\pi(j)} + x_{j+1} y_{\pi(j+1)} > x_j y_{\pi(j+1)} + x_{j+1} y_{\pi(j)}, \tag{8.6.36}$$

so switching these two y's reduces the sum (35). A finite number of these transpositions leads to the reversing permutation, which therefore minimizes.

This same argument yields the following generalization. Instead of (35) we minimize

$$f(x_1, y_{\pi(1)}) + \ldots + f(x_n, y_{\pi(n)}). \qquad (8.6.37)\ \ (37)$$

Then, if f has non-negative cross-differences, the reversing permutation minimizes (37). If f has positive cross-differences, this minimizer is unique. Inequality (36) is replaced by the cross-difference inequality. Note that (35) is the special case of (37) where f is the product function: $f(x,y) = xy$).

As one suspected, there is an allotment-assignment problem lurking about. Indeed, let $S = Q = \{1,\ldots,n\}$, $\Sigma_s = \Sigma_q =$ all subsets, $\alpha = \beta =$ ~~enumeration~~ counting measure, $h(i) = x_i$, and $w(j) = y_j$ for $i,j = 1,\ldots,n$, and f be as in (37). The resulting allotment-assignment problem reads as follows.

Minimize the sum of the n^2 terms

$$f(x_i, y_j)\nu_{ij} \qquad (8.6.38)\ \ (38)$$

(i,j ranging independently over $\{1,\ldots,n\}$), over all non-negative (n,n) matrices (ν_{ij}) whose rows and columns all sum to 1 ("bistochastic matrices").

The values (37) are embedded among the values (38). Specifically, for feasible matrices consisting of just 0s and 1s ("permutation matrices"), (38) reduces to (37). If f has positive cross-differences, the unique optimal solution is the "weight-falloff" measure given by: $\nu_{ij} = 1$ if $i + j = n + 1$, $\nu_{ij} = 0$ otherwise. This solution corresponds to the reversing permutation above.

The allotment-assignment problem just constructed is a special case of the <u>assignment</u> problem of ordinary linear programming. ~~As is well-known,~~ this in turn is a special case of the transportation problem. The positive cross-differences property of the objective function enables us to read off the optimal solution at sight.

A

8.7. The Thünen System as a Social Equilibrium: Formal Theory

A social equilibrium is a system involving several agents (with possibly conflicting preferences) such that no agent can take any action that ~~which~~ improves the situation according to his own preferences. An example is the real-eatate market of chapter 6, in which agents acquire regions of Space (or Space-Time). Here each agent has a preference ordering over pairs consisting of the region acquired and the cost of acquiring it. Equilibrium consists of a pattern of real-estate values (represented by a measure over S or $S \times T$) and a partition of S (or $S \times T$) among the agents, such that no agent can improve his position by switching to another region under the existing pattern of prices.

The equilibrium of this section is similar to that of the real-estate market, but goes a step deeper. Namely, we assume that people acquire land ~~in order~~ to operate land uses on it. The decision problem facing each agent is accordingly more complicated: He must decide not only what land to acquire but what to do with it — i.e. ~~that is~~, what the land-use assignment is to be. The separate decisions of the various agents then result in a pattern of land uses over the whole system. We will show that, under certain mild assumptions, this overall land-use assignment is a Thünen system, in that it satisfies the measurable weight-falloff condition.

Now ~~for~~ consider the formal model. We are given the measure space (S,Σ_s,α), where S is physical Space, and α is the areal measure on its ~~sigma~~ σ-field Σ_s. The measurable function $h{:}S \to$ reals gives the distance of locations from the nucleus. Also given is the measurable space of land uses, (Q,Σ_q), ~~together~~ with the measurable weight function $w{:}Q \to$ reals. Area, distance, and weight are all "ideal" and may be considerably distorted from the corresponding physical magnitudes (recall section 8.2). Specifically, they have the following properties. Let ν be a measure on $(S \times Q, \Sigma_s \times \Sigma_q)$ representing a certain land-use assignment: $\nu(E \times F)$ is the (ideal) acreage in region E devoted to land uses of types F. Then α measures the capacity of regions to accommodate land uses, in the sense that any feasible ν must satisfy

$$\nu(E \times Q) \leq \alpha(E) \qquad (8.7.1)\ ~~(1)~~$$

for all regions E. Also, the transportation cost incurred in region E by assignment ν is

$$\int_{E\times Q} f(h(s),w(q))\,\nu(ds,dq). \qquad (8.7.2)\ ~~(2)~~$$

fl

Here f:reals^2 → reals is a given measurable function; one case we have mentioned often is the product function, $f(x,y) = xy$.

9a 29-7 We now introduce a countable (possibly finite) number of agents, labeled $n = 1,2,\ldots$. At time zero, when the system starts up, there is a big real-estate auction ~~which~~ that leads to a measurable partition of Space, S, among the agents. S_n is the region falling under the control of agent n. Upon acquiring S_n, agent n chooses an assignment ν_n, which is a measure over $S_n \times Q$. The only constraint on ν_n is that it satisfy (1) for all subregions E of S_n. (Agent n will also have a budget constraint, but we suppose that this is reflected indirectly in his preference ordering to be discussed below, and therefore need not be taken into account explicitly). The several assignments ν_n on $S_n \times Q$, $n = 1,2,\ldots$, then yield by direct summation an overall assignment ν on $S \times Q$.

~~There are~~ two kinds of costs are incurred by agent n. The first is transportation cost, which is given by (2) with $E = S_n$, $\nu = \nu_n$. The second is land cost. We suppose ~~that~~ the real-estate market leads to a system of land values ~~which~~ that is represented by a measure (or perhaps a signed measure) μ over Space. The net cost of land to agent n is

$$\mu(S_n). \qquad (8.7.3)\ ~~(3)~~$$

($\mu(S_n)$ is actually the opportunity cost of land, in the sense that, even if agent n uses his own land, and therefore pays no rent, he still loses the opportunity to rent or sell to someone else.) Total cost incurred by him is the sum of land cost and transportation cost.

We now come to the structure of agents' preferences. The idea is to make assumptions ~~which~~ that are very weak, yet ~~which~~ lead to substantive conclusions. Let β_n be the right marginal of assignment ν_n; ~~That is,~~ i.e.

$$\beta_n(F) = \nu_n(S_n \times F)$$

for all $F \in \Sigma_q$. It is natural to call β_n the *allotment* corresponding to assignment ν_n. We now assume that the preference ordering of agent n satisfies the following condition: If two different actions yield the same allotment, the first is at least as preferred as the second iff the cost incurred under the first does not exceed the cost incurred under the second. Symbolically this may be written

$$(\beta,\ c_1) \succsim_n (\beta,\ c_2) \text{ iff } c_1 \le c_2. \qquad (8.7.4)\ (4)$$

Here c_i is the cost incurred under action i, $i = 1,2$, β being the common allotment. No assumption is made concerning preferences among actions leading to different β's; these need not even be comparable. Also preferences may vary in an arbitrary way from agent to agent, except that they all satisfy (4).

The rationale for (4) is worth examining in some detail. At first glance (4) appears to be so weak as to border on tautology. It states that, ~~other things being equal~~ ceteris paribus, more money is preferred to less. This condition would appear to be universally satisfied, except possibly for the small minority for whom poverty is a virtue, and even these have the option of throwing excess money away. In particular, (4) should be distinguished from the much stronger condition of profit maximization (or ~~rather~~ cost minimization in this case). Cost minimization entails indifference between two actions yielding the same cost, whereas (4) states nothing concerning actions that ~~which~~ yield different allotments.

(4) does, however, carry an ~~some~~ implicit substantive assumptions: ~~First of all, agent n is indifferent to what anyone else does, since he considers only his own allotment and costs. In reality, of course, agent n would not be indifferent to the existence of public services, or of retail stores at which he can buy, etc.; nor would he be indifferent to neighborhood effects, such as pollution, emanating from land uses on parcels near his own land. Secondly,~~ Agent n is indifferent to all

no ¶

aspects of his own land-use assignment other than allotment and cost. These other aspects include layout, shape of parcel, whether his land is in one piece or fragmented, etc. From one point of view this assumption is not implausible. Consider, for example, an agent contemplating two actions, both of which yield an identical allotment of one acre devoted to a certain residential land use, three acres devoted to his various business activities, two acres for recreation, etc. ~~and so on~~. The agent's life-style will be the same in both cases, as will his ~~and his~~

income-expenditure pattern ~~will be the same in both cases~~ (except for transportation and land costs). The only difference lies in the spatial distribution of these activities, and why should this concern him? The only reason is that transportation plus land costs may vary with the spatial arrangement, and this factor is already taken into account in (4).

start 29-8

Thus the argument for (4) reduces, roughly, to the following: If one can do the same thing in region A or region B, one should be indifferent to location, except for cost. This is fine except for one difficulty, and that is to make sure that all spatially-varying costs are counted in. Now the transport cost formula (2) embodies the basic Thünen assumption that all trips involve the nucleus either as origin or destination (or are two-leg trips passing through the nucleus). In reality there are also "local" trips, such as walking from room to room in one's house. The cost of these local trips will be higher in narrow or fragmented parcels than in "chunky" parcels. The postulated preference orders are unrealistic to the extent that these cost components are omitted. (A related omission is that of possible neighborhood effects among an agent's own land uses).

Given the preference order of agent n on land-use assignments and regions, this induces an indirect preference order on regions alone, ~~namely,~~ viz. $E_1 > E_2$ by agent n iff for any land-use assignment on region E_2 there is an assignment on E_1 at least as preferred. This invites comparison with the preference orderings in the real-estate market of chapter 6, which are also over regions. In general, the approach of this section does not lead to an additive utility function of the type discussed in chapter 6.

We ~~Let us~~ now gather up the strands of this discussion in the following definition.

D

Definition: A social equilibrium for the system (S,Σ_s,α), (Q,Σ_q), h, w, f, and preference orders $\succcurlyeq_n$, $n = 1,2,\ldots$, consists of

18 (i) a signed measure μ on (S,Σ_s) ("land values");

(ii) a measurable partition (S_n), $n = 1,2,\ldots$, of S; ~~and~~

(iii) measures ν_n on $S_n \times Q$, $n = 1,2,\ldots$;

18 such that the capacity constraint (1) is satisfied by the ν_n, and ~~such that,~~ for each agent n, total cost is finite, and the pair

$$\left[\nu_n'' \;,\; \int_{S_n\times Q} f(h(s),w(q))\,\nu_n(ds,dq) + \mu(S_n)\right] \qquad (8.7.5)\ ~~(5)~~$$

cannot be surpassed in the agent's ~~his~~ preference order by substitution of any other feasible region and assignment for S_n and ν_n.

Here the orderings $\succcurlyeq_n$, the partition elements S_n, and the measures ν_n all have the same countable index set, $n = 1,2,\ldots$; and $\nu_n'' = \beta_n$ is the right marginal of ν_n. In short, each agent finds that his choice (S_n,ν_n) yields an allotment-cost pair (5) that ~~which~~ is unsurpassed in the set of options available ~~to him~~.

Note that, unlike the situation in the allotment-assignment problem, the allotments β_n are not given in advance, but are to be chosen by the agents. ~~Note~~ also, ~~that~~ μ is ~~sigma~~ σ-finite, since land cost to each agent is finite.

Before launching into details, ~~let us~~ we mention an alternative (less realistic) way of defining social equilibrium for the system above. The definition just given requires <u>exclusive</u> control of land, in the sense that Space is partitioned, and agent n alone decides on the land-use assignment in the region S_n ~~which~~ that he acquires. Suppose we now drop this requirement, and allow <u>joint</u> control of land. Agent n ~~has to~~ must choose merely a land-use assignment ν_n, which is now a measure on $S \times Q$. ~~Note that~~ The universe set for ν_n is the entire product space for each agent. There is no longer a need for agents to partition Space among themselves into proprietary regions.

To spell out this alternative model we must specify constraints and costs. Let ν be the sum of all individual assignments:

$$\nu = \nu_1 + \nu_2 + \nu_3 + \ldots \qquad (8.7.6)$$

Then the areal capacity constraint (1) remains valid, with ν given by (6). That is, the area required by all agents together must not exceed the area available, so there is one global constraint. Transportation cost incurred by agent n is simply (2), with $E = S$ and $\nu = \nu_n$.

gal 9

Land cost is a bit more complicated. Just as above, a land-value (signed) measure μ over Space arises in the real-estate market. This cost is then prorated among the agents in proportion to the area they require in various regions. To be precise, we have for each n,

$$\nu_n' \leq \nu' \leq \alpha$$

from (6) and (1), where the prime denotes the left marginal. Hence ν_n' is absolutely continuous with respect to ν'; and since α (hence ν') is ~~sigma~~ σ-finite, the Radon-Nikodym theorem asserts the existence of a function g_n:S → reals such that

$$\nu_n' = \int g_n \, d\nu'. \qquad \text{(8.7.7)} \ \cancel{(7)}$$

We have $0 \leq g_n(s) \leq 1$ for ν'-almost-all locations s, and g_n represents the pro rata share of land of agent n. The land (opportunity) cost for agent n is then

$$\int_S g_n \, d\mu. \qquad \text{(8.7.8)} \ \cancel{(8)}$$

Under exclusive control, (8) reduces to (3).[30] ~~A similar integral arises from the pro rata initial holdings of agent n, and the difference of these two integrals gives land outlay, as in (3).[29]~~

start

Everything else is the same as in the "exclusive control" model. Preferences still satisfy (4), where β_n is the right marginal of ν_n. Social equilibrium is given by a land-value signed measure μ, and assignments (ν_n), $n = 1,2,\ldots$, such that the capacity constraint (1) is satisfied and ~~such that~~ the allotment-cost combination yielded by ν_n is unsurpassed in the preference order of agent n, for all n.

There is a simple relation between the "exclusive" and "non-exclusive" concepts of social equilibrium. Namely, the former may be thought of as the special case of the latter in which the additional requirement is imposed that the assignments of any two agents be mutually singular in the following strong sense: There is a measurable partition, $(S_1, S_2, \dots)$ of S such that $\nu_n[(S \setminus S_n) \times Q] = 0$, for all n. To be precise, we identify these measures ν_n (which all have universe set $S \times Q$) with their restrictions to $S_n \times Q$, $n = 1,2,\dots$. These restricted measures are of the form used in the "exclusive control" model. It is now easy to check that the "non-exclusive" versions of areal constraints, transportation and land costs reduce to the "exclusive" versions.

Given the land-use assignments $\nu_1, \nu_2, \dots$, the overall pattern of land uses ν in the "non-exclusive control" model is given by summation (6). If the mutual singularity condition just mentioned obtains, and we re-interpret ν_n to refer to the restriction of agent n's assignment to $S_n \times Q$, then the same ν is obtained by direct summation:[31]

$$\nu = \nu_1 \oplus \nu_2 \oplus \dots \qquad (8.7.9)\ \cancel{(9)}$$

~~As mentioned above,~~ the "non-exclusive control" model is in most respects less realistic than the "exclusive control" model. Joint control of land is relatively rare. Even rarer is a free market in partial shares of land: Generally, ~~one~~ a person

~~has to~~ must buy (or rent) a parcel completely or not at all. Finally, where such "fractional" markets do exist, it is unreasonable to ignore neighborhood effects from one agent's activities to another, since these activities are not merely "adjacent" but thoroughly "mixed" with each other. There are, however, some situations in which something resembling joint control is in effect: easements, joint usage of public facilities such as roads or parks, and perhaps land control by organizations.

In the following we ~~shall~~ restrict attention to the "exclusive control" model. For the record, however, we state that the "non-exclusive control" equilibrium can also be shown to have an overall land-use pattern satisfying the weight-falloff condition. The argument for this is similar to the one given below; it is even a bit simpler, and does not need the "non-atomic" assumption (10). (For notational convenience we replace boldface Q by ordinary Q in the following formal discussion.)

Theorem: Given ~~sigma~~ σ-finite measure space (S,Σ_s,α), measurable space (Q,Σ_q), measurable functions $h{:}S \to$ reals, $w{:}Q \to$ reals, $f{:}\text{reals}^2 \to$ reals, with f having positive cross-differences, and

$$\alpha\{s \mid h(s) = x\} = 0 \qquad \text{(8.7.10)} \quad \cancel{(10)}$$

for all real x. Let $(\succ_n)$, $n = 1,2,\ldots$, be a countable family of partial orders each satisfying (4).

Let there be a social equilibrium for this system consisting of signed measure μ on S, measurable partition (S_n), $n = 1,2,\ldots$, of S, and feasible measures ν_n on $S_n \times Q$, $n = 1,2,\ldots$ Let $f(h(\cdot), w(\cdot))$ be half-bounded on each set $S_n \times Q$.

Then the overall assignment ν (given by (9)) satisfies the (measurable) weight-falloff condition (with respect to h, w).

Proof: First we show that each individual ν_n, considered in isolation as a measure on $S_n \times Q$, must satisfy measurable weight-falloff. Indeed, consider any other measure $\tilde{\nu}_n$ on $S_n \times Q$ ~~which~~ that has the same left and right marginals as ν_n. Since $\tilde{\nu}_n$ satisfies the areal capacity constraint (1), it is feasible for agent n, and therefore cannot surpass ν_n in his preference ordering. Now ν_n and $\tilde{\nu}_n$ have the same allotment; they also occupy the same region, S_n, hence incur the same land cost. It follows that the transportation cost (2) incurred under ν_n cannot exceed the transportation cost incurred under $\tilde{\nu}_n$.

~~In other words,~~ Thus ν_n is unsurpassed for the allotment-assignment problem on $S_n \times Q$ defined by its marginals and by f, h and w. Since $f(h(\cdot), w(\cdot))$ is half-bounded on this set, and f has positive cross-differences, it follows (~~page~~ 000) that ν_n does indeed satisfy measurable weight-falloff.

Now suppose the overall assignment ν violates measurable weight-falloff, so that there are measurable sets, G_1, $G_2 \subseteq$

$S \times Q$, of positive ν-measure, with G_1 southwest of G_2. This means that S and Q can each be split into two measurable sets, S^1, S^2, and Q^1, Q^2, respectively, such that

$$h(s_1) < h(s_2),$$ (8.7.11) ~~(11)~~

~~and~~

$$w(q_1) < w(q_2),$$ (8.7.12) (12)

for all $s_i \in S^i$, $q_i \in Q^i$, $i = 1,2$, and such that

$$G_i \subseteq S^i \times Q^i,$$ (8.7.13) ~~(13)~~

$i = 1,2$.

~~A~~ Relation (13) implies that $S^1 \times Q^1$ and $S^2 \times Q^2$ both have positive ν-measure. Since the number of agents is countable, there must be some agent, n_1, such that $\nu[(S_{n_1} \cap S^1) \times Q^1]$ is positive. Furthermore, since $\nu' \leq \alpha$ is ~~sigma~~ σ-finite, ~~there is~~ a region E_1 exist such that $\nu[(E_1 \cap S_{n_1} \cap S^1) \times Q^1]$ is positive and <u>finite</u>. Similarly, there is an agent n_2 and region E_2 such that $\nu[(E_2 \cap S_{n_2} \cap S^2) \times Q^2]$ is positive and finite.

extra lead to clear

The two agents, n_1 and n_2, must be <u>distinct</u>, for if not, the measurable weight-falloff condition would be violated within one individual realm, contrary to our finding above.

Now consider the two functions g_i:reals $\to$ reals, $i = 1,2$, given by

12 pt brackets + parens

$$g_1(x_1) = \nu\left[\left(\{s \mid h(s) > x_1\} \cap E_1 \cap S_{n_1} \cap S^1\right) \times Q^1\right]$$

~~and~~ (8.7.14) (14)

$$g_2(x_2) = \nu\left[\left(\{s \mid h(s) < x_2\} \cap E_2 \cap S_{n_2} \cap S^2\right) \times Q^2\right].$$

fl l

These functions are non-negative but not identically zero, monotone (non-increasing for g_1 and non-decreasing for g_2), approach zero asymptotically; and are — because of (10) — continuous. Hence ~~there exist~~ numbers $x_1°$, $x_2°$ exist such that

$$0 < g_1(x_1°) = g_2(x_2°) < \min\left[\lim_{x_1 \to -\infty} g_1(x_1),\ \lim_{x_2 \to \infty} g_2(x_2)\right]. \qquad (8.7.15)\ \cancel{(15)}$$

12pt brackets please visi

Now define the regions H^1, H^2 by

$$H^1 = \{s \mid h(s) > x_1°\} \cap E_1 \cap S_{n_1} \cap S^1,$$

$$H^2 = \{s \mid h(s) < x_2°\} \cap E_2 \cap S_{n_2} \cap S^2. \qquad (8.7.16)\ \cancel{(16)}$$

From (15) we then have

$$\infty > \nu(H^1 \times Q^1) = \nu(H^2 \times Q^2) > 0. \qquad (8.7.17)\ \cancel{(17)}$$

Let c be the common value of $\nu(H^i \times Q^i)$. Next, define region J^1 by

$$J^1 = \{s \mid h(s) \leq x_1°\} \cap E_1 \cap S_{n_1} \cap S^1. \qquad (8.7.18)\ \cancel{(18)}$$

From the right-hand inequality in (15), it follows that $\nu(J^1 \times Q^1)$ is positive. Now the two sets, $J^1 \times Q^1$ and $H^1 \times Q^2$, stand in a southwest-northeast relation, by (12), (16) and (18). Furthermore, both H^1 and J^1 are subsets of S_{n_1}. It follows that

$$\nu(H^1 \times Q^2) = 0, \qquad (8.7.19)\ \cancel{(19)}$$

otherwise ~~else~~ the measurable weight-falloff condition would be violated within the individual realm of agent n_1. A similar argument for agent n_2 shows that

$$\nu(H^2 \times Q^1) = 0. \qquad (8.7.20) \quad \cancel{(20)}$$

Relations (17), (19) and (20) are what is needed for the rest of this proof.

In what follows, the notation ν_G stands for the truncation of ν to the set G; i.e. ~~that is~~, for $G, K \in \Sigma_s \times \Sigma_q$, we have

$$\nu_G(K) = \nu(G \cap K). \qquad (8.7.21) \quad \cancel{(21)}$$

Now consider the following policy changes. Agent n_1, instead of taking S_{n_1} as his region of control, chooses $(S_{n_1} \backslash H^1) \cup H^2$. That is, he relinquishes control of region H^1 and acquires region H^2. Agent n_2 makes just the opposite switch, so that his region of control becomes $(S_{n_2} \backslash H^2) \cup H^1$. Also, the original pattern of land uses on $H^1 \cup H^2$ is altered. Specifically, the following signed measure is added to the original overall land-use assignment ν:

$$\Delta\nu = -\nu_{H^1 \times Q^1} - \nu_{H^2 \times Q^2} + \frac{1}{c}\left[(\nu_{H^2 \times Q^2})' \times (\nu_{H^1 \times Q^1})''\right] \qquad (8.7.22) \quad \cancel{(22)}$$

$$+ \frac{1}{c}\left[(\nu_{H^1 \times Q^1})' \times (\nu_{H^2 \times Q^2})''\right].$$

or all on one line if possible

(To explain: The fourth term in (22), ~~for example~~ e.g., is the product measure of the left marginal of $\nu_{H^1 \times Q^1}$ by the right marginal of $\nu_{H^2 \times Q^2}$, divided by c, the common value in (17).) On H^1, because of (19), the first term of (22) knocks out the original assignment; the fourth term is what replaces it; similarly, on H^2, because of (20), the second term knocks out the original assignment, and the third term is what replaces it.

~~Let us~~ We first check this new assignment for feasibility. All measures in (22) are finite, and it is clear that $\nu + \Delta\nu$ is non-negative. Using (17), ~~one~~ we ~~verifies~~ verify that the left marginal of $\Delta\nu$ is identically zero. Hence the areal capacity constraint (1) remains satisfied for $\nu + \Delta\nu$. This establishes feasibility.

Next, ~~one~~ we ~~verifies~~ verify that the right marginal of the first plus third ~~tarm~~ term of (22) is identically zero. ~~Now~~ These two terms give precisely the change from the original assignment of agent n_1. Thus the allotment attained by agent n_1 is exactly the same as with his original assignment. Since the original assignment of agent n_1 is unsurpassed in his preference ordering, it follows from (4) that the change in total cost must be non-negative. Thus

$$\int_{S\times Q} f(h(\cdot),w(\cdot))\,\Big|\, d\left[\frac{1}{c}\left((\nu_{H^2\times Q^2})' \times (\nu_{H^1\times Q^1})''\right) - \nu_{H^1\times Q^1}\right] + \mu(H^2) - \mu(H^1) \geq 0. \qquad (8.7.23)\ \cancel{(23)}$$

The integral in (23) is the change in the transport cost incurred by agent n_1. The other two terms yield the change in his land cost, incurred by acquiring region H^2 and relinquishing H^1. ~~(Note that all terms in (23) are bounded, so the expression is well-defined.)~~ (Finiteness of costs in the social equilibrium, together with the half-boundedness condition, insure that expression (23) is well-defined).

The same argument applied to agent n_2 yields (23) with the superscripts "1" and "2" interchanged. Adding (23) to the corresponding inequality for n_2, the land-value terms drop out, and we obtain

$$\int_{S\times Q} f(h(\cdot),w(\cdot))\,\Big|\, d(\Delta\nu) \geq 0. \qquad (8.7.24)\ \cancel{(24)}$$

But (24) is false. To show this we use the argument already employed ~~on page above,~~ in (5.17) through (5.19) ~~(17) through (19) of section 5~~. The applicability of this argument follows from the observations that f has positive cross-differences, that $H^1 \times Q^1$ is southwest of $H^2 \times Q^2$ (from (11) and (12)), and that all measures in (22) have the same value on $S \times Q$, ~~namely~~ viz. c. $\Delta\nu$ of (22) then has the same form as ν in (5.11) ~~(11) of section 5~~, except for the inessential factor c: Just let G_i in (5.10) ~~(10) of section 5~~ be $H^i \times Q^i$ of the present argument, $i = 1,2$. It follows that the integral in (24) must be negative. This contradiction shows that ν cannot violate measurable weight-falloff.

10/12 X 27 TR q1=2ems

Note that, while this theorem severely constrains the pattern of land uses, it says nothing about land users. The individual regions S_n controlled by the various agents might have any irregular shapes, be fragmented, intermixed, etc. But the kinds of land uses ~~running~~ at a site will depend (as ~~so~~ far as their ideal weights are concerned) only on the ideal distance of that site, not on the agent who controls it ~~that site~~. Smith will run a land use on one of his plots similar in weight to that run by Jones on an adjacent plot, and dissimilar to the land use run by Smith on another of his plots at a different ideal distance from the nucleus.

Before going on, we ~~let us~~ mention a ~~two~~ generalization ~~generalisations~~ of the preceding theorem, that is of theoretical but little practical interest. no ¶ ~~First.~~ The premise that land value, μ, must be bounded on each region S_n can be dropped from the definition of social equilibrium. Instead, μ can be any σ ~~sigma~~-finite signed measure on S (or even a pseudomeasure) without invalidating the conclusion that assignment ν satisfies the measurable weight-falloff condition.[32] ~~Second, the premise that $f(h(\cdot),w(\cdot))$ be bounded on each set $S_n \times Q$ can be weakened to its being half-bounded on the union of any two of these sets. We omit proofs of these statements.~~

We ~~Let us~~ now turn from the study of land uses to land values. Our aim is to show that land-value density is essentially a (left) half-potential for the allotment-assignment problem associated with the foregoing social equilibrium. But before doing so, we must face up to certain conceptual difficulties ~~which~~ that we managed to evade in the preceding proof.

The first difficulty concerns the assumption of perfect competition embodied in our definition of social equilibrium. To be precise, we assume that agents can freely acquire or dispose of land at the fixed prices given by the signed measure μ. In reality, search and bargaining problems arise between the transacting agents, which are unlikely to be captured by an additive set function such as μ. (Transfer of land involves the displacement of one pattern of land uses by another on this land, hence changes in the allotments of the several agents; in general, the monetary compensations for these changes will not take a form ~~which~~ that is additive over regions.)

As a rule, the more agents there are, the more satisfactory the competitive assumption becomes. But even with a countably infinite number of agents the difficulty does not vanish, since these agents will still hold positive acreages.[33]

The second difficulty involves a slightly paradoxical strengthening of the social equilibrium concept. We start with an analogy. Suppose one has a divisionalized firm, the various divisions trading $\boxed{\text{both with}}$ the outside world and ~~with~~ each other. A possible rule of operation is that each division offer the same prices both to outsiders and sister-divisions, and that each accepts the best offer, ignoring the affiliation status of trading partners. In short, the divisions act as if they were independent firms. Now consider agent n holding region S_n. Partitioning this region into $S_{n1},\ldots,S_{nk}$, we may think of the land uses on each piece S_{ni} as being run by a "division" of agent n. We now make the "independent firms" assumption about the behavior of these divisions. To be precise, the "custodian" of region S_{ni} has the option of buying a piece of the territory of his alter-ego, the "custodian" of S_{nj} $(j \neq i)$, at the going market price, and displacing the latter's land uses on the transferred region. The change in costs resulting from this transaction is to be computed just as if S_{nj} were controlled by an agent other than n — ~~that is~~ i.e., one takes account neither of the land uses ousted from S_{nj}, nor of the compensation received by the custodian of this region.

This rule is rather awkward to justify if taken literally, and the correct interpretation seems to be as follows. The land-value measure μ plays a double rôle. Externally, between agents, it functions as a market price; internally, within an agent's territory, it functions as a "shadow" price, equilibrium under the above rule being a necessary condition for the optimal assignment of land uses. If there is just one agent in the entire system, μ plays exclusively a "shadow" rôle. As the number of agents rises, the "shadow" rôle shrinks and the "market" rôle expands, but the former does not disappear even for a countable infinity of agents. (It *would* disappear in the "measure space of agents" model.)

930-2

In the preceding proof we avoided these conceptual difficulties by centering the argument around a swap of territories ~~which~~ that reduced combined transportation costs. In the following proof we are not so lucky. The special "independent firms" rule shows up in the proof below by allowing the argument to go through even if the agent "acquires" some of his own land.

We give Now for the details. The assignment ν on $S \times Q$ induces a measure λ on the plane via the mapping $(s,q) \rightarrow (h(s),w(q))$. Let λ and the function f:reals^2 $\rightarrow$ reals be given.

D

Definition: The measurable function p:reals $\rightarrow$ extended reals is a *left half-potential for λ-almost-everywhere* (with respect to f) iff there is a real Borel set F with $\lambda'(F) = 0$ such that $p(x)$ is finite for $x \notin F$ and

$$p(x_1) + f(x_1,y_2) \geq p(x_2) + f(x_2,y_2) \qquad \text{(8.7.25)} \; \cancel{(25)}$$

for all real numbers x_1, x_2, y_2 such that x_1, $x_2 \notin F$ and (x_2,y_2) is a point of support for λ.

Here λ' is the left marginal of λ. If $F = \emptyset$, this reduces to the ordinary definition of left half-potential, as in (5.67) ~~(67) of section 5~~.

In (27) below, $\alpha - \nu'$ is the *areal excess capacity measure*. Since both ν' (the left marginal of ν) and α may be infinite measures, subtraction must be understood in the sense of ~~chapter 3,~~ section 3 1.

Theorem: Let overall assignment measure ν on $S \times Q$, land value signed measure μ on S, and regions S_n, $n = 1,2,\ldots$, constitute a social equilibrium for the system defined by (S,Σ_s,α), (Q,Σ_q), $h:S \to$ reals, $w:Q \to$ reals, f:reals$^2 \to$ reals, $\succ_n$, $n = 1,2,\ldots$, as in the preceding theorem.

Here α is ~~sigma~~ σ-finite, h, w, and f are measurable, and, as always, ν satisfies (1), $\succ_n$ satisfies (4), all n, and all agents have finite costs. In addition, assume that h and α satisfy (10), that f is continuous with positive cross-differences, ~~and~~ that $f(x,y)$ is strictly increasing in x for all numbers y in the range of w, and that $f(h(\cdot), w(\cdot))$ is half-bounded on each set $S_n \times Q$.

Then:

(i) There is a non-negative, non-increasing function p:reals $\to$ extended reals, ~~which~~ that is a left half-potential for λ almost everywhere (with respect to f) — (here λ is the plane measure induced from ν by $(s,q) \to (h(s),w(q))$) — and for which

$$\mu = \int (p\circ h)\,d\alpha. \qquad (8.7.26)\ \cancel{(26)}$$

(Hence $\mu \geq 0$).

(ii) There is an extended real number, x_o, such that

$$(\alpha-\nu')\{s|h(s) \leq x_o\} = 0, \qquad (8.7.27)\ \cancel{(27)}$$

$$\nu'\{s|h(s) \geq x_o\} = 0, \qquad (8.7.28)\ \cancel{(28)}$$

and

$$\mu\{s|h(s) \geq x_o\} = 0. \qquad (8.7.29)\ \cancel{(29)}$$

Proof: (i) First we show that $\mu \geq 0$. Suppose $\mu(E)$ were negative for some region E. Since μ is ~~sigma~~ s-finite, there is a subregion $F \subseteq E$ for which $\mu(F)$ is negative and finite. Choose any agent n, and let him acquire region F and leave it vacant. Then his land cost falls, while his transport cost and his allotment remain the same; this contradicts the fact that he was in equilibrium to begin with. Hence $\mu \geq 0$. (Note that the agent may be "acquiring" land from himself, in accordance with our discussion above.)

Next ~~let us~~ we show that μ is absolutely continuous with respect to ν'. Let $\mu(E) > 0$ for some region E. Then $\mu(E \cap S_n) > 0$ for some n, since the sets S_n countably partition S. If $\nu'(E \cap S_n) = 0$, agent n could simply divest himself of $E \cap S_n$, reducing his land cost while leaving his transport cost and allotment unchanged. This contradicts the un-surpassedness of his initial position. Hence $\nu'(E \cap S_n) > 0$, so that $\nu'(E) > 0$. Thus $\mu \ll \nu'$.

$\nu' \leq \alpha$ is ~~sigma~~ s-finite. By the Radon-Nikodym theorem, it now follows that ~~there~~ exists a function $\tilde{p}: S \to$ reals such that

$$\mu = \int \tilde{p}\, d\nu'. \qquad (30) \quad (8.7.30)$$

Since μ is a measure, $\tilde{p}$ may be chosen to be non-negative.

Now consider the mapping from S to the plane given by $s \to (h(s), \tilde{p}(s))$. This induces a measure, ρ, on the plane from the measure ν' on S. We now show that ρ satisfies the weight-falloff condition. Suppose, on the contrary, that there are two measurable sets, F_1 and F_2, in the plane, of positive ρ-measure, with F_1 southwest of F_2. The inverse images of these sets are regions, G_1 and G_2, respectively, of positive ν'-measure, such that ~~there~~ exist numbers x_o, z_o, satisfying

$$h(s_1) \leq x_o \leq h(s_2), \tag{8.7.31}$$

$$\tilde{p}(s_1) \leq z_o \leq \tilde{p}(s_2), \tag{8.7.32}$$

for all $s_1 \in G_1$, $s_2 \in G_2$. Furthermore, at least one of the inequalities in (31), and at least one in (32), can be chosen strict. (This follows from the definition of "southwest".)

Now (10) holds with ν' in place of α. Using an argument similar to (14)–(17), we can find two subregions, H_1 and H_2, of G_1 and G_2, respectively, such that H_2 is contained in the realm of some one agent: $H_2 \subseteq S_n$, say, such that ~~f and~~ μ is bounded on $H_1 \cup H_2$, and ~~such that~~

$$\infty > \nu'(H_1) = \nu'(H_2) > 0. \tag{8.7.33}$$

Let c be the common value in (33).

Consider now the following changes in the action of agent n. He relinquishes region H_2 and acquires H_1. The corresponding change in his assignment is

$$\Delta\nu = -\nu_{H_2 \times Q} + \frac{1}{c}\left[(\nu_{H_1 \times Q})' \times (\nu_{H_2 \times Q})''\right]. \tag{8.7.34}$$

12pt brackets

The notation is as in (21) and (22). The first term in (34) knocks out his original assignment on H_2; the second term is the assignment he places on the newly acquired region, H_1.

(Displacement of the original assignment on H_1 does not appear in (34), since it does not appertain to agent n. As discussed above, this displacement is to be ignored even if $H_1 \cap S_n$ is non-empty).

Let us test this for feasibility. It suffices to take regions $E \subseteq H_1$ and to check that

$$\Delta\nu(E \times Q) = \frac{1}{c}\nu(E \times Q)\nu(H_2 \times Q) = \nu'(E)$$

$\nu(E\times Q)\,\nu(H_2\times Q)/c$

does not exceed $\alpha(E)$, as indeed it does not, since the original assignment was feasible.

The right marginal of $\Delta\nu$ is identically zero, so that the allotment of agent n is unchanged.

The change in land cost is given by

$$\mu(H_1) - \mu(H_2) = \int_{H_1} \tilde{p}\, d\nu' - \int_{H_2} \tilde{p}\, d\nu',$$

from (30). But

$$\int_{H_1} \tilde{p}\, d\nu' \leq z_o c \leq \int_{H_2} \tilde{p}\, d\nu', \qquad (8.7.35)\ ~~(35)~~$$

from (32) and (33). Furthermore, one of the inequalities in (32), hence in (35), is strict. It follows that land cost has been reduced.

The change in transport cost is given by

$$\int_{S\times Q} f(h(s),w(q))\,\Delta\nu(ds,dq). \qquad (8.7.36)\ ~~(36)~~$$

30-4

But

$$\int_{S\times Q} f(h(s),w(q))\,\nu_{H_2\times Q}(ds,dq) \geq \int_{S\times Q} f(x_o,w(q))\,\nu_{H_2\times Q}(ds,dq)$$

$$= \int_{Q} f(x_o,w(q))\,(\nu_{H_2\times Q})''(dq) \qquad (8.7.37)\ \text{~~(37)~~}$$

WILL NOT FIT WITH ALIGNMENT -LEE

$$\geq \int_{S\times Q} f(h(s),w(q))\left[\frac{1}{c}(\nu_{H_1\times Q})'(ds) \times (\nu_{H_2\times Q})''(dq)\right].$$

The first integral in (37), being within the original realm of agent n, is finite. Hence all the integrals in (37) are well-defined. The first inequality in (37) arises from the right side of (31) and the fact that $f(\cdot,y)$ is increasing for y in the range of w; the equality arises from the induced integrals theorem, using the projection $(s,q) \to q$. The measure $(\nu_{H_2\times Q})''$ is the right marginal of both measures in (34). Hence, passing over to the second measure in (34), and using the left side of (31) and the increasingness of $f(\cdot,y)$ once again, we obtain the last inequality of (37). Furthermore, at least one of the inequalities in (31), hence in (37), is strict.

But (36) equals the last integral of (37) minus the first integral of (37). Thus transport cost is also reduced. The change in *total* cost is then *negative*. This improvement contradicts the fact that agent n is in equilibrium.

Thus the measure ρ does indeed satisfy the weight-falloff condition. Expressing this in topological form, no two points of support for ρ stand in a southwest-northeast relation.

18 Now let X be the set of real numbers x with the following property: There is exactly one number y such that (x,y) is a point of support for ρ. Define the function $p: X \to$ reals by letting $p(x)$ be the unique number for which $(x, p(x))$ supports ρ.

X is a Borel set. ~~For $X = X_1 \setminus X_2$, where~~ To show this, let

$$X_i = \{x \mid (x,y) \text{ supports } \rho \text{ for at least } i \text{ distinct numbers } y\},$$

$i = 1,2$. Set X_1 is the left projection of the support of ρ. The support is a closed set, hence a countable union of bounded closed sets; X_1 itself is therefore a countable union of closed sets, hence Borel. X_2 is also a Borel set, since it is countable. To see this, associate with each $x \in X_2$ a rational number between y' and y'', where (x,y'), (x,y'') are distinct points of support for ρ. These rationals are distinct, by weight-falloff. Since the rationals are countable, so is X_2. But $X = X_1 \setminus X_2$. Thus X is Borel.

We now show that X includes "almost all" the real numbers in the following sense. Let ρ' be the left marginal of ρ. Note that ρ' is also the measure on the real line induced by h from ν' on S. For, letting E be a Borel set, we have

$$\rho'(E) = \rho(E \times \text{reals}) = \nu'\{s \mid h(s) \in E\}. \tag{8.7.38}$$

We now assert that

$$\rho'(\text{reals} \setminus X) = 0. \tag{8.7.39}$$

First ~~of all~~, $\rho'(\text{reals}\setminus X_1) = 0$, since ρ is zero on the complement of its support. Secondly, $\rho'(X_2) = 0$, since X_2 is countable and ρ' is zero on each singleton, by (10) and (38). This proves (39).

Now extend p from domain X to the entire real line as follows: ~~For $x \geq \sup X$, let $p(x) = 0$. For $x < \sup X$, let~~ $p(x)$ ~~be~~ is the supremum of zero and the values $p(z)$ for $z \in X$, $z \geq x$. This extended function will also be denoted by p. The original p is non-negative and non-increasing, and one easily verifies that the extended p retains these properties.

Next we show that the composite function $p \circ h$ is equal to $\tilde{p}$, ν'-almost everywhere on S. Indeed, consider the region H given by

$$\{s \mid h(s) \in X\} \cap \{s \mid (h(s), \tilde{p}(s)) \text{ supports } \rho\}. \qquad (8.7.40)\ \ ~~(40)~~$$

For any site $s \in H$, there is exactly one number y such that $(h(s), y)$ supports ρ, so that $p(h(s)) = y = \tilde{p}(s)$. Thus $\tilde{p}$ and $p \circ h$ coincide on H. The complement of the first region in (40) has ν'-measure zero, by (38) and (39). The complement of the second region in (40) also has ν'-measure zero, since ρ is zero on the complement of its support. Thus $\tilde{p} = p \circ h$ almost everywhere.

¶ It follows from (30) that

$$\mu = \int (p \circ h)\, d\nu'. \qquad (8.7.41)\ \ ~~(41)~~$$

Now consider the mapping from $S \times Q$ to the plane, given by $(s,q) \to (h(s),w(q))$. This induces a measure, λ, on the plane from the measure ν on $S \times Q$. Since ν is a social equilibrium, the preceding theorem implies that it, hence also λ, satisfies weight-falloff. Expressed in topological form, this means that no two points of support for λ stand in a southwest-northeast relation.

Let Z be the set of real numbers z with the following property: There is exactly one number y such that (z,y) supports λ. Using the same arguments for Z and λ that we used above for X and ρ, we conclude that Z is a Borel set, and that

$$\lambda'(\text{reals} \setminus Z) = 0, \qquad (8.7.42)$$

where λ' is the left marginal of λ. We also have $\lambda' = \rho'$, since, for Borel sets E,

$$\lambda'(E) = \lambda(E \times \text{reals}) = \nu\{(s,q) \mid h(s) \in E\} = \nu'\{s \mid h(s) \in E\}.$$

(cf. (38)).

We now verify that the function p is a left half-potential for λ almost everywhere. Specifically, we show that (25) holds for all x_1, $x_2 \in X \cap Z$, and all real y_2 such that (x_2, y_2) supports λ. Note that, from (39) and (42), the complement of $X \cap Z$ has λ'-measure zero, so the "almost everywhere" condition is met.

Suppose (25) were false, so that ~~there~~ exist numbers x_1, $x_2 \in X \cap Z$, and y_2 real for which (x_2, y_2) supports λ, and

$$p(x_2) + f(x_2, y_2) > p(x_1) + f(x_1, y_2). \qquad (8.7.43)\ \cancel{(43)}$$

Suppose (which may not be the case) there were an increasing sequence of numbers, $x^1 < x^2 < \ldots$, belonging to $X \cap Z$, with limit x_2. For each n, let y^n be the number such that (x^n, y^n) supports λ. Then $y^1 \geq y^2 \geq \ldots$, and these numbers are bounded below by y_2. Hence $\lim y^n$ exists as $n \to \infty$, and is finite. Since the support of λ is closed, it follows that $(x_2, \lim y^n)$ belongs to it. Hence

$$\lim_{n\to\infty} y^n = y_2, \qquad (8.7.44)\ \cancel{(44)}$$

since there is exactly one point of the form (x_2, y) supporting λ. A similar argument establishes (44) in the case of a decreasing sequence, $x^1 > x^2 > \ldots$, belonging to $X \cap Z$, with limit x_2.

Furthermore, the same argument, applied to the support of ρ instead of λ, yields

set

165 24

$$\lim_{n\to\infty} p(x^n) = p(x_2),$$

for any monotone sequence (x^n) belonging to $X \cap Z$ with limit x_2. This argument applies also to any number $x \in X \cap Z$, in particular to $x = x_1$. Thus p, restricted to $X \cap Z$, is continuous.

30-6

Now let $b > 0$ be the difference between the left and right ~~hand~~ sides of (43). Choose $\varepsilon > 0$ so that all the following relations are satisfied for $i = 1$ and $i = 2$:

$$|p(x) - p(x_i)| < b/4, \text{ for all } x \in X \cap Z \text{ such that } |x - x_i| < \varepsilon, \cancel{\text{and}} \qquad \text{(8.7.45)} \; \cancel{(45)}$$

$$|f(x,y) - f(x_i,y_2)| < b/4, \text{ for all } x, y \text{ such that } |x - x_i| < \varepsilon \text{ and } |y - y_2| < \varepsilon. \qquad \text{(8.7.46)} \; \cancel{(46)}$$

The existence of such an ε follows from the continuity of p at points x_1 and x_2, and the continuity of f at points (x_1,y_2) and (x_2,y_2). Next, choose a $\delta > 0$ so that

$$\lambda\left\{(x,y)\,\middle|\,|x - x_2| < \delta,\ |y - y_2| \geq \varepsilon\right\} = 0. \qquad \text{(8.7.47)} \; (47)$$

The existence of such a δ follows from (44). For δ can be chosen so small that for any $x \in X \cap Z$ such that $|x - x_2| < \delta$, the y such that (x,y) supports λ is within ε of y_2. The set in (47) will therefore own ~~have~~ no points of support, hence has measure zero.

Having chosen ε, then δ, consider the two regions G_i given by

$$\{s \mid |h(s) - x_i| < \min(\varepsilon,\delta)\},$$

$i = 1,2$. Both these regions have positive ν'-measure, since there are numbers y_1, y_2 such that (x_i,y_i) supports λ, $i = 1,2$. (The following argument does not require that G_1, G_2 (or H_1, H_2 for that matter) be disjoint).

one line if possible or break for

set first and last larger as shown

12 pt brackets a vertical bar

12 pt braces

set first and last larger

Just as above, we can now find two subregions, H_1 and H_2, with $H_i \subseteq G_i$, $i = 1,2$, such that H_2 is contained in the realm of some one agent: $H_2 \subseteq S_n$, and ~~such that~~ (33) holds.

Continuing just as above, we contemplate the policy change for agent n in which he relinquishes region H_2 and acquires H_1. The corresponding change in his assignment is again given by $\Delta\nu$ of (34).

As above, this change is feasible and leaves his allotment unchanged. The reduction in his land cost arising from the relinquishing of region H_2 is given by

centered degree

$$\int_{H_2} (p \circ h)\, d\nu' = \int_{H_2 \times Q} p(h(s))\, \nu(ds,dq). \qquad (8.7.48)\ \cancel{(48)}$$

The left integral in (48) comes from (41). The equality arises from the induced integrals theorem.

CONT. 30-6 X 300 LEAD TO 90!

Adding to this the reduction of transport cost on H_2, we find the reduction of *total* cost arising from the relinquishing of region H_2 to be

12 pt brackets

$$\int_{H_2 \times Q} \Big(p(h(s)) + f(h(s),w(q))\Big)\, \nu(ds,dq). \qquad (8.7.49)\ \cancel{(49)}$$

We now claim that the integral (49) exceeds cm, where c is the common value $\nu'(H_1) = \nu'(H_2)$, and m is the average of the two sides of (43). To see this, split the set $H_2 \times Q$ into two pieces: $H_2 \times L$, and $H_2 \times (Q \setminus L)$, where

{ | } larger where marked

12 pt braces & vertical bar

$$L = \{q \mid |w(q) - y_2| < \varepsilon\}.$$

Now $\nu(H_2 \times (Q\backslash L)) = 0$, from (47). However ~~On the other hand~~, for all points of $H_2 \times L$ (except for a ν-null set), the integrand in (49) exceeds m. (The exceptional points (s,q) are among those for which $h(s) \notin X \cap Z$. These have measure zero, by (39) and (42).) For (45) and (46) imply that the integrand value differs from the left side of (43) by less than $b/2$, hence it never gets halfway toward the right side of (43). It follows that (49) exceeds $m\cdot\nu(H_2 \times Q) = mc$.

raised ⊙

Next, consider the increase in costs arising from the acquisition of region H_1. For land cost this is

centered again

$$\int_{H_1}(p\circ h)\,d\nu' = \int_{H_1\times Q} p(h(s))\left[\frac{1}{c}(\nu_{H_1\times Q})'(ds) \times (\nu_{H_2\times Q})''(dq)\right]. \qquad \text{(8.7.50)} \; ~~(50)~~$$

The left integral in (50) comes from (41). The equality arises from the induced integrals theorem, on noting that the left marginal of the bracketed measure in (50) coincides with ν' truncated to H_1.

g 30-7

Adding to this the increase in transport cost, we find the increase in __total__ cost arising from this acquisition to be

$$\int_{H_1\times Q}[p(h(s))+f(h(s),w(q))]\left[\frac{1}{c}(\nu_{H_1\times Q})'(ds)\times(\nu_{H_2\times Q})''(dq)\right]. \qquad \text{(8.7.51)} \; ~~(51)~~$$

The bracketed measure in (51) has value zero on $H_1 \times (Q\backslash L)$, from (47). However ~~On the other hand~~, for all points of $H_1 \times L$ (except for a null set as above), the integrand in (51) is less than m. For (45) and (46) imply that the integrand differs from the right side of (43) by less than $b/2$, hence it never gets halfway toward the left side ~~of (43)~~. It follows that (51) is less than

$$m\left[\frac{1}{c}(\nu_{H_1\times Q})'(H_1)\cdot(\nu_{H_2\times Q})''(L)\right] = m\cdot\frac{1}{c}\cdot c\cdot c = mc,$$

from (47), since

$$(\nu_{H_2\times Q})''(L) = \nu(H_2\times L) = \nu(H_2\times Q) = c.$$

Thus (49) exceeds (51). Hence the net increase in costs, which is (51) minus (49), is negative. Since his allotment remains the same, this contradicts the premise that agent n was initially in an unsurpassed position. Thus (43) must have been false, and we have established that p is an almost everywhere left half-potential for λ.

This completes part (i) of the theorem, except for the fact that ν' appears in (41) while α appears in (26). The proof of part (ii)~~, which we come to now,~~ will justify the last step of replacing ν' by α.

(ii) Let x_1 be the infimum of all numbers x satisfying

$$\nu'\{s|h(s) > x\} = 0, \qquad (8.7.52)\ ~~(52)~~$$

and let x_2 be the supremum of all numbers x satisfying

$$(\alpha-\nu')\{s|h(s) < x\} = 0. \qquad (8.7.53)\ ~~(53)~~$$

pel

~~There are~~ three cases follow:

If $x_1 < x_2$, let x_o be any number satisfying $x_1 < x_o < x_2$. Equations (27) and (28) then follow at once, and (29) follows from (28), since, as shown above, μ is absolutely continuous with respect to ν'.

If $x_1 = x_2$, let x_o be their common value. (52) then holds for $x = x_o$, because the set $\{s|h(s) > x_o\}$ is the limit of an increasing sequence of sets of ν'-measure zero. Similarly, (53) holds for $x = x_o$. Also, $\alpha-\nu'$ and ν' are both zero on the set $\{s|h(s) = x_o\}$, from (10). This again establishes (27) and (28), hence (29).

To finish the proof, we need only show that the last possibility, $x_1 > x_2$, leads to a contradiction. Choose a number x_o such that $x_1 > x_o > x_2$. Then

$$\nu'\{s|h(s) > x_o\} > 0, \qquad \cancel{(54)} \tag{8.7.54}$$

from (52), and

$$(\alpha-\nu')\{s|h(s) < x_o\} > 0, \qquad \cancel{(55)} \tag{8.7.55}$$

from (53). From (55), ~~there~~ exists an agent n for which

$$(\alpha-\nu')[S_n \cap \{s|h(s) < x_o\}] > 0. \qquad \cancel{(56)} \tag{8.7.56}$$

For the time being, we also assume that

$$\nu'[S_n \cap \{s|h(s) > x_o\}] > 0. \qquad \cancel{(57)} \tag{8.7.57}$$

Using an argument similar to (14) ~~through~~ (17), we can find regions H_1 and H_2, contained in the sets of (56) and (57), respectively, such that

$$\infty > (\alpha-\nu')(H_1) \geq \nu'(H_2) > 0.$$

Now consider the following change in policy of agent n. He relinquishes region H_2 and increases his assignment on H_1. Specifically, the change in his assignment, $\Delta\nu$, is given by

$$-\nu_{H_2\times Q} + \frac{(\alpha-\nu')_{H_1} \times (\nu_{H_2\times Q})''}{(\alpha-\nu')(H_1)}. \qquad (8.7.58)\ \cancel{(58)}$$

930-8

The left measure in (58) knocks out the original assignment on H_2. Note that the original assignment on H_1 is not displaced, but added to by the right measure in (58). The excess capacity on H_1 allows the areal constraint (1) to remain inviolate, as one verifies. Hence the changed policy is feasible.

The right marginal of $\Delta\nu$ is identically zero, so that the allotment of agent n remains unchanged. The land cost is reduced, if anything, since he divests himself of H_2. (There is no change in land cost for H_1, since he controls this region to begin with.) Transportation cost must fall, by the argument of (37). (There is an obvious change of notation for the last integral of (37), and both inequalities there are now strict.) Hence total costs fall, and agent n has improved his position. This contradicts the fact that he is initially in equilibrium. Thus (57) must be false.

break at 30-7

(divests)

This same argument proves even more. Start with (56) and (57), with *any* number x substituted for x_o (the same x in both (56) and (57)). The same contradiction arises, so (56) and (57) cannot hold simultaneously for *any* x in place of x_o.

It follows from this that a (possibly infinite) number x_{oo} exists such that

$$\nu'[S_n \cap \{ s | h(s) \geq x_{oo}\}] = 0 \qquad (8.7.59) \quad \cancel{(59)}$$

and

$$(\alpha-\nu')[S_n \cap \{s | h(s) \leq x_{oo}\}] = 0. \qquad (8.7.60) \quad \cancel{(60)}$$

(To see this, apply the argument at the beginning of part (*ii*) to the region S_n, rather than to all of S.)

From (60) and (56) we obtain

$$(\alpha-\nu')[S_n \cap \{s | x_o > h(s) > x_{oo}\}] > 0. \qquad (8.7.61) \quad \cancel{(61)}$$

$\alpha-\nu'$ can be replaced by α in (61), because of (59). Also,

(vertical)

$$\mu[S_n \cap \{s | x_o > h(s) > x_{oo}\}] = 0, \qquad (8.7.62) \quad \cancel{(62)}$$

from (59).

Now, from (54), ~~there~~ exists an agent n^* for which

$$\nu'[S_{n^*} \cap \{s | h(s) > x_o\}] > 0. \qquad (8.7.63) \quad \cancel{(63)}$$

Arguing in the usual way, we can find regions H and H^*, contained in the sets of (61) and (63), respectively, such that

$$\infty > \alpha(H) \geq \nu'(H^*) > 0.$$

Consider the following policy changes by agent n^*. He relinquishes region H^* and acquires H, making the change of assignment $\Delta\nu$ given by

$$-\nu_{H^*\times Q} + \frac{\alpha_H \times (\nu_{H^*\times Q})''}{\alpha(H)}.$$

One verifies the feasibility of this change, and the fact that the allotment for agent n^* remains the same.

The change in land cost is nonpositive, because H is free, by (62). The change in transport cost is negative, by the argument of (37) (with H^* in place of H_2, and obvious changes of notation in the last integral of (37); both inequalities in (37) are strict). Thus agent n^* has reduced his total costs and improved his position, contradicting the fact that n^* is initially in equilibrium.

H star, H sub two

This shows that $x_1 > x_2$ is false, where x_1, x_2 are defined by (52) and (53). The proof of part (ii) is now complete.

#

18 Finally, ~~let us~~ we show that

131 20 83

$$\left|\int_E\right| (p\circ h)\, d\nu' = \left|\int_E\right| (p\circ h)\, d\alpha \qquad \text{(8.7.64)} \; \cancel{(64)}$$

centered degree

for all regions E. In conjunction with (41), this will prove (26). First, α and ν' coincide on all subregions of

$\{s|h(s) \leq x_o\}$, from (27). Hence (64) is true for any E contained in this set. Secondly, for $E \subset \{s|h(s) > x_o\}$, ~~we have μ(E) = 0, from (29). Hence~~ the left side of (64) is zero, from (~~41~~ 28). As for the right side, we first note that

$$\sup X \leq x_o. \qquad \cancel{(65)}\ (8.7.65)$$

g 30-9

To prove (65), let $x \in X$. Then (x,y) supports ρ for some number y, which implies that

$$\nu'\{s| \quad |h(s) - x| < \varepsilon\} > 0, \qquad \cancel{(66)}\ (8.7.66)$$

12 pt braces + vertical bar

{|} larger where marked

for all $\varepsilon > 0$, from (38). If x exceeded x_o, (66) would contradict (28). This proves (65). But $p(x) = 0$ for all $x > \sup X$, so that $p(h(s)) = 0$ for all $s \in E$. Thus the right side of (64) is also zero, and (64) is again true. Finally, any region E splits into sets of these two types, so (64) is true in general. This completes the proof of (26). ~~|||~~ □

\#

These results are worth contemplating. The number x_o is the "natural radius" of the Thünen system, and corresponds to the classical "margin of cultivation". Beyond x_o land is permanently vacant and free, while land closer than x_o is filled to capacity. (But "filled" with lower and lower density uses as one moves outward~~, of course.~~) Note that x_o can equal $+\infty$, indicating an infinite Thünen system with no boundary.

The natural radius is "almost" unique in the following sense: x_{oo} is another natural radius iff

$$\alpha\{s|h(s) \text{ is between } x_o \text{ and } x_{oo}\} = 0. \qquad \cancel{(67)}\ (8.7.67)$$

fl

(This statement ~~easily~~ follows from (27) and (28).) In the classical Thünen situation, for example, where S is the plane, h is Euclidean distance from the nucleus, and α is ~~Euclidean~~ ordinary area, (67) implies $x_{oo} = x_o$, so the natural radius is indeed unique.

The fact that p is an (almost everywhere) left half-potential yields ~~a great deal of~~ much information about p (and thus about land values, via (26)), for the results of section ~~5~~ 8 apply. To see what is involved, define k:reals $\to$ extended reals by

$$k(y) = \inf\{p(x) + f(x,y) \mid x \in X \cap Z\}, \qquad (8.7.68)\ \cancel{(68)}$$

where X and Z are defined as in the preceding proof. This implies that

$$p(x) = \sup\{k(y) - f(x,y) \mid y \text{ real}\}, \qquad (8.7.69)\ \cancel{(69)}$$

for all $x \in X \cap Z$. To prove (69), first note that (68) implies

$$p(x) \geq k(y) - f(x,y) \qquad (8.7.70)\ \cancel{(70)}$$

for all $x \in X \cap Z$, all real y. Let $x_2 \in X \cap Z$ be given, and let y_2 be the (unique) number such that (x_2, y_2) supports λ. From (25) it follows that the infimum in (68) for $y = y_2$ is attained at $x = x_2$. Hence (70) is satisfied with equality at (x_2, y_2), so that (69) is proved for $x = x_2$.

The pair (p,k) have most of the characteristics of a topological potential for λ (relative to f). But k, unlike p, has no simple intuitive interpretation. (One may strain a bit and interpret $k(y)$ as the maximum cost per unit area ~~on~~ an agent operating a land use of weight y would ~~put up with~~ tolerate).

(69) may now be applied as in section 8. For example, if $f(\cdot, y)$ is a concave function for each y, then p, as the supremum of convex functions, is convex. More precisely, if $x_1, x_2, x_3 \in (X \cap Z)$ and $x_2 = \theta x_1 + (1-\theta)x_3$, where θ lies between 0 and 1, then

$$p(x_2) \leq \theta p(x_1) + (1-\theta)p(x_3). \tag{71}$$

(8.7.71)

(Since (71) may not apply to all triples x_1, x_2, x_3, the stated condition might better be called "convexity almost everywhere").

B

A Simplified Approach

The foregoing results concerning the structure of a Thünen social equilibrium, while rather striking, are also rather complicated to derive. At the same time, the underlying arguments are basically quite simple, though this fact is obscured by the requirements of rigor. It seems worthwhile, then, to present an alternative approach ~~which~~ that is heuristic and informal. The sacrifice of rigor extends not only to the reasoning but even to the concepts themselves. Thus the notion of "running a land use at a location" is inherently vague, and must be translated into measure-theoretic language to attain clarity. But to carry out all the clarifications needed would simply land us back in the complexities of the preceding proofs. Let us therefore plunge ahead boldly.[34]

g30-10

Let there be a social equilibrium, with agent n_i running land use q_i at site s_i, $i = 1,2$. The weight of land use q_i is $w(q_i) = y_i$, and the distance of site s_i from the nucleus is $h(s_i) = x_i$, $i = 1,2$.

We assume as usual that the transport cost incurred per acre at distance x for weight y is given by $f(x,y)$, where f has positive cross-differences. Also assume that there is a land-value density function $\tilde{p}(s)$.

Now agent n_1 has the option of relinquishing an acre of land "at" site s_1, acquiring an acre of land "at" site s_2, and switching his land use q_1 from s_1 to s_2. This does not change his allotment; hence, since he is initially in equilibrium, this change cannot reduce his total costs. Thus

$$\tilde{p}(s_1) + f(x_1,y_1) \leq \tilde{p}(s_2) + f(x_2,y_1). \qquad \text{(8.7.72)} \; \cancel{(72)}$$

The left side of (72) is approximately the total cost incurred by agent n_1 from the running of one acre of land use q_1 at site s_1. The right side is about what his total costs would be from running q_1 instead on an acre at s_2. A similar argument applies to agent n_2, who has the option of switching his land use q_2 from site s_2 to s_1. This yields

$$\tilde{p}(s_2) + f(x_2,y_2) \leq \tilde{p}(s_1) + f(x_1,y_2) \qquad \text{(8.7.73)} \; \cancel{(73)}$$

(which is just (72) with subscripts interchanged).

These are the key relations. Adding (72) and (73), $\tilde{p}$ drops out and we obtain

$$f(x_1,y_1) + f(x_2,y_2) \le f(x_1,y_2) + f(x_2,y_1). \qquad (8.7.74)\ \cancel{(74)}$$

Now if $x_1 < x_2$, we cannot have $y_1 < y_2$, for otherwise (74) would contradict the fact that f has positive cross-differences; ~~That is~~ i.e., it cannot happen that a lighter land use is "at" a nearer site while a heavier use is "at" a more distant site. This may be taken as a heuristic characterization of the weight-falloff condition.

Next, take the special case where $x_1 = x_2$. Then (72) and (73) ~~then~~ yield $\tilde{p}(s_1) \le \tilde{p}(s_2) \le \tilde{p}(s_1)$, so that $\tilde{p}(s_1) = \tilde{p}(s_2)$; ~~That is~~ i.e., whenever two sites have the same ideal distance, they have the same land-value density. It follows that $\tilde{p}$ may be written as a composite function, $p \circ h$. (73) then reduces to

$$p(x_2) + f(x_2,y_2) \le p(x_1) + f(x_1,y_2). \qquad (8.7.75)\ \cancel{(75)}$$

This is a heuristic form of the condition (25) that p be a left half-potential for λ, the plane measure induced by the mapping $(s,q) \to (h(s),w(q))$ from the assignment ν.

From (75) one can derive several properties of p from corresponding properties of f. The two most interesting are monotonicity and convexity:

Let $f(\cdot,y)$ be strictly increasing for each y, and choose $x_1 < x_2$. Assume there is a location s_2 at distance x_2 from the nucleus, with a land use q_2 in operation "at" s_2; let y_2

be the weight of q_2. Then (75) is satisfied for this x_1, x_2, y_2. Hence

$$p(x_1) - p(x_2) \geq f(x_2,y_2) - f(x_1,y_2) > 0,$$

so that p is strictly decreasing.

Let $f(\cdot,y)$ be a concave function for each y, and let x_1, x_2, x_3 satisfy: $x_2 = \theta x_1 + (1-\theta)x_3$, where θ is a number between 0 and 1. Assume as above that there is a location s_2 at distance x_2 with a land use q_2 of weight y_2 operating "at" s_2. (75) is then satisfied for this x_1, x_2, y_2, and we also have

$$p(x_2) + f(x_2,y_2) \leq p(x_3) + f(x_3,y_2), \qquad \cancel{(76)}\ (8.7.76)$$

upon substituting x_3 for x_1 in (75). Now multiply (75) by θ, (76) by $(1-\theta)$, and add. After rearrangement we obtain

$$\theta p(x_1)+(1-\theta)p(x_3)-p(x_2) \geq f(x_2,y_2)-\theta f(x_1,y_2)-(1-\theta)f(x_3,y_2). \qquad \cancel{(77)}\ (8.7.77)$$

But the right side of (77) is non-negative, since $f(\cdot,y_2)$ is concave. Hence p is a convex function. (This argument can be spruced up to provide an alternative rigorous derivation of the "almost everywhere convexity" property (71)).

Finally, we indicate briefly how one goes from (75) to the important line integral representation of p (cf. ~~(83) of section 5)~~ (5.83)):

30

216

$$p(z_1) - p(z_2) = \oint_{z_{\cancel{2}1}}^{z_{\cancel{1}2}} D_1 f(x,y)\,dx. \qquad \text{(8.7.78)} \quad \cancel{(78)}$$

Here $D_1 f(x,y)$ is the partial derivative of f with respect to its first argument, and the integral is taken along the line of support in the plane, connecting the points of support for λ. ~~(see pages above).~~ We may assume that $z_1 < z_2$. Choose a sequence $x_0 < x_1 < \ldots < x_n$, with $x_0 = z_1$, $x_n = z_2$, and let y_i be a number such that (x_i, y_i) is on the line of support, $i = 0, \ldots, n$. We then obtain

zero

$$\begin{aligned} \big(f(x_i, y_i) - f(x_{i-1}, y_i)\big) &\le \big(p(x_{i-1}) - p(x_i)\big) \\ &\le \big(f(x_i, y_{i-1}) - f(x_{i-1}, y_{i-1})\big) \end{aligned} \qquad \text{(8.7.79)} \quad (79)$$

12 pt brackets were blue parens are marked

Ed is this correct to reg page paths

for $i = 1, \ldots, n$. The left inequality in (79) arises from substituting the triple of numbers (x_{i-1}, x_i, y_i) for (x_1, x_2, y_2) in (75); the right inequality in (79) arises from substituting the triple (x_i, x_{i-1}, y_{i-1}) for (x_1, x_2, y_2) in (75).

Now add the inequalities (79) over $i = 1, \ldots, n$. The p terms in the middle add to $p(x_0) - p(x_n)$, which is the left side of (78). Also, ~~On the other hand, with the proper continuity assumptions,~~ assuming continuity of $D_1 f$, both the left and right ~~hand~~ f-differences in (79) can be approximated by $(x_i - x_{i-1}) D_1 f(x_i, y_i)$. Adding this over i yields essentially a Riemann sum for the line integral in (78). ~~Under the proper assumptions,~~ A limit argument then establishes (78).

zero

$D_1 f$ one

9a 31-1

(A)

8.8. The Thünen System as a Social Equilibrium: Discussion

10/12 x 27 TR ф=2ems

One striking fact about the model just developed is its great generality, or — what is the same thing — the weakness of the assumptions from which we start. ~~To~~ We recapitulate the main ideas. First, area, distance and weight are all "ideal", which allows great flexibility in incorporating geographic and institutional irregularities. Second, no assumption is made concerning the region controlled by any one agent. It may consist of many scattered parcels, near and far, of irregular shapes. Third, the only assumption concerning preferences is (7.4) ~~(4) of section 7~~, which may be roughly stated as: "Other things being equal, lower costs are preferred to higher."

This generality of preferences is especially important for the real-estate market, such as we consider ~~are dealing with~~ here. In industry analysis one deals with businessmen engaged in similar activities and ~~one~~ has some basis for assuming similar motivations (typically ~~one~~ assuming they all maximize income). But in the real-estate market we have religious, commercial, governmental, residential, etc., land users all participating. Thus philanthropists and robber barons, bureaucrats and businessmen, are all competing cheek-by-jowl in the same market, and one can hardly assume uniformity of preferences among these agents. (These remarks assume there are no zoning restrictions. With a zoning law the market splits to some extent into "noncompeting groups", but still with considerable diversity within these groups).

The two other main assumptions~~, it is true,~~ are rather more restrictive. These are: the absence of any constraint on land-use assignments other than the areal capacity constraint, (7.1) ~~(1) of section 7~~, and the characteristic form for transportation costs, (7.2) ~~(2) of section 7~~ (with f having positive cross-differences). In the ~~next~~ section 8.9 we ~~shall~~ argue that these assumptions apply to a broader range of cases than might be expected at first glance.

We ~~Let us~~ now examine the foregoing Thünen system from the point of view of social welfare. Is the social equilibrium at which the market arrives optimal in some sense? We know that the equilibrium assignment ν satisfies the measurable weight-falloff condition, and that it is therefore optimal for the allotment-assignment problem defined by its marginals and the transport cost integral, (7.2) ~~(2) of section 7~~ (assuming certain other weak premises)~~; see page above)~~. That is, ν minimizes total transport costs over the set of all measures on $S \times Q$ having left- and right-marginals ν' and ν'', respectively. This has been vaguely recognized (but never really proved or stated clearly) by several writers who claim that land uses arrange themselves to minimize the "friction of space".

But does this property imply social optimality? According to Alonso it does not, because the reduction of transport cost is just one desideratum, which must be balanced against others, such as freedom from congestion and the quality of life.[35] This statement is ~~of course~~ correct; however, it is also irrelevant for the question under discussion. ~~The reason is that~~ because ν minimizes transport cost over a set of measures with the <u>same</u> right marginal, ν'', and this constraint has the effect of <u>holding</u> <u>constant</u> all the other desiderata.

To spell this out, let ν_n on universe set $S_n \times Q$ be the assignment of agent n, $n = 1,2,\ldots$. We then have

$$\nu'' = \nu_1'' + \nu_2'' + \ldots,$$

where ν_n'' is the right marginal of ν_n. Consider now any alternative system of assignments, $\tilde{\nu}_n$, $n = 1,2,\ldots$, yielding overall assignment $\tilde{\nu}$ by direct summation, such that these satisfy

$$\tilde{\nu}_n'' = \nu_n'', \tag{8.8.1}$$

for all $n = 1,2,\ldots$, and

$$\tilde{\nu}' = \nu'. \tag{8.8.2}$$

Adding (1) over all n, we obtain

$$\tilde{\nu}'' = \nu''. \tag{8.8.3}$$

(2) and (3) then imply that transport cost under $\tilde{\nu}$ is at least as large as under ν. At the same time, (1) indicates that the mode of life of all agents is the same under these two situations ~~(see pages above)~~.

Since all other desiderata are held constant, it would seem that minimization of transport cost is a <u>necessary</u> condition for social optimality. As in so many other contexts, it appears that here too the competitive market solution has socially desirable features.

5a 2 However, the same inefficiencies that crop up in competitive solutions also arise here. Transport cost ~~here~~ is the sum of all individual costs, as personally assessed by each separate agent. Now in transportation particularly ~~there are~~ all sorts of costs are imposed on other agents, which are not assessed against the traveler or shipper, and would not be counted in his private costs. These "external" costs include delays and crowding for other travelers, increased risks of accident for them, pollution and noise, etc.[36] Second, the market implicitly weights the preferences of different agents by treating all dollars equally, no matter who spends them; one may decide to reject this weighting on ethical grounds. In either of these cases, the fact that the market minimizes "total transport costs" loses its normative significance.[37]

36

37

~~Let us~~ We now turn to the structural implications of the model. Since the Thünen social equilibrium ~~turns out to be~~ is an optimal solution to the allotment-assignment problem, the extensive discussion of section ~~6~~ 8 applies to it. Thus we get Thünen "rings" of land uses of decreasing weight as we go outward from the nucleus. The "rings" tend to be elongated along major transport arteries -- leading to "urban sprawl" -- and to form disconnected subcenters around points of access to limited-access transportation systems (highway interchanges, airports, railway stations~~, etc.~~). This latter process yields a "central place" hierarchy of centers and land uses,[38] the "higher-order" activities being those ~~which~~ that are heavier.

Now consider the weight-falloff property more closely. First, ~~let us~~ note what it does not state. It says nothing about how much land is to be allotted to various uses, or even whether a given land use will appear in the system at all.[39] It only states that if two land uses do appear, the heavier will be closer to the nucleus (or, at worst, equidistant from it). ~~Now~~ Cities exhibit certain broad regularities in the ordering of their land uses by distance, and it is enlightening to compare these regularities with the predictions of the weight-falloff property.

Consider multiple-story structures. ~~As discussed above, page ______,~~ There is a general (but not perfect) tendency for land uses involving many stories to be heavier than land uses involving few stories. We would expect, then, the skyline of a city to get lower as one moves away from the nucleus (~~ie.~~ i.e., the CBD), -- as ~~of course~~ indeed it does.

The exceptions to this rule are no less enlightening than the uses ~~which~~ that conform to it. The exceptions are there for most part thanks to political intervention ~~which~~ that consciously sets itself against the natural tendency of the market — either by direct public ownership, in the case of streets, parks, etc., or by laws forcing private owners to limit the coverage of their lots and the size and bulk of their buildings. The case of the transportation system is particularly interesting. There is some tendency for transportation to conform to the general pattern, with elevated highways, subways, etc., in downtown areas; but ~~clearly~~ the stacking of transportation surfaces is not carried ~~less~~ as far ~~than~~ as the stacking of other kinds of land uses. Actually, transportation is not a land use in the sense in which we are using the term in the Thünen model. Here "land use" is by definition sedentary. Thus there is no strong reason to expect that transportation would conform to the general weight-falloff pattern.

The analysis just given identifies "Space" with the surface of the Earth and considers vertically stacked floors to be parts of one land use. ~~There is~~ an alternative view ~~which~~ identifies "Space" with horizontal surfaces of support, including the separate floors of a multiple-story structure (see Section 8.4). ~~(see page above)~~. On this view, each story supports a separate land use. This alternative view yields some additional predictions. The ideal distance of a site from the nucleus now includes not only the cost of surface transportation, but also vertical transportation costs if the site is not at ground level. Thus ideal distance rises with each successively higher story (successively lower in the case of subterranean structures).

Applying the model to this situation, one expects a progressive lightening of land uses as one moves vertically away from ground level, together with a fall in the value of floor space — just as if one had moved horizontally further from the nucleus. (A rough analogy would be the climatic zones on a mountain at the equator; the zones ~~one~~ passes d through going upward approximate those ~~one~~ passes d through going poleward.)

Are these predictions borne out? In general, rental value falls as one goes from the ground floor to the second or third floors, and this pattern conforms to expectations. But (if there is an elevator) it levels out and perhaps rises for higher floors.[40] To put these facts in perspective, ~~one should~~ note the following points. First, successive floors are not perfect substitutes for each other — either technically (in terms of the weight of machinery, vibration, etc., they can sustain) or psychologically: (Higher floors have cleaner air, less street noise, commanding views, and the general quality of "upmanship" ~~which~~ that leads people to climb mountains, sit on raised platforms, and wear elevator shoes). This lack of homogeneity violates one of the conditions of the model, and helps explain the rise in rentals on upper floors.

Second, the "ideal" height of even the tallest existing skyscraper is rather small: just a few minutes of travel time. If it took, say, a half-~~an~~-hour to get to the top we would begin to find some palpable differentiation among stories. ~~Note~~ also ~~that~~ much of the incremental cost is incurred in the first floor or two (in the form of waiting time or stair-climbing), which helps explain the rapid initial fall of rent. (The well-known specialization of ground floors in retail trade seems to be due to their high visibility from the street, a factor ~~which~~ that again violates one of the conditions of the model. This also contributes to the initial fall in rent.)

Let us now turn to land speculation ~~- that is~~ i.e., the policy of delaying the onset of a land use to some point in the future. As discussed ~~above, page~~ in Section 8.4, this delayed activity is a land use in its own right, the forward time-displacement of the original activity. ~~It was noted that,~~ in general, a forward time-displaced land use is lighter than the original. Hence, if use q and its displacement q' both appear in a Thünen system, q' will locate further from the nucleus than q, by the weight-falloff condition.

We would expect, then, a general tendency for land uses to become more delayed in onset as ~~one~~ we move~~s~~ outward from the nucleus. The initial fallow period, before any imports or exports arise from the site, should become longer with distance. This is observed, and is ~~in fact~~ the phenomenon of suburbanization.

At the same time, ~~there are~~ certain exceptions to this rule ~~which~~ are also implied by the weight-falloff conditions.

Let q_1 be a land use and q_1' a forward time-displacement of q_1. While q_1' is lighter than q_1, it may well be heavier than another land use q_2 whose onset occurs before that of q_1'. In this case, if q_1' and q_2 both appear in the system, q_1' will locate closer to the nucleus than ~~will~~ q_2. There will then be an interval of time during which a nearer region (that occupied by q_1') will remain vacant while a further region (occupied by q_2) will be busily importing and exporting. This is the phenomenon of leapfrogging, which has been widely observed and commented on. ~~The point to~~ This note ~~is~~ that it does not constitute an "intensity-reversal" contradicting the weight-falloff condition, if "intensity" is correctly measured as an integral over the entire time-horizon of the system, and not as a short-run indicator.[41]

The typical land uses of the central business district — ~~such as~~ (front-office and professional activities, advertising, finance, ~~government~~ department stores, theaters, night clubs, and hotels) — should be the heaviest in the system, since they are the most central. Their weight arises from their eminent "stackability" into multiple stories and the high densities of people with relatively high-cost time involved in them.[42]

At the other extreme, typically suburban land uses, such as golf courses and cemeteries, shade off finally into agriculture. These are all very light compared with most urban land uses, and the pattern again conforms to weight-falloff.

Manufacturing is found at all distances, and this reflects the heterogeneous character of this class of land uses. Suburban manufacturing is typically a one-story affair, with large areas devoted to parking, storage, and landscaped grounds, while central manufacturing tends to occur in lofts with a densely-crowded labor force. ~~Note, by the way, that~~ the typical designation of certain manufacturing activities as "light" or "heavy" has nothing to do with their ideal weights. The "heavy" activity of steelmaking, for example, is much lighter in ideal weight than the "light" activity of apparel manufacture; ~~and~~ the much more centralized location of the latter activity bears this out.

In general, the higher the average population density associated with a given land use, the heavier the weight of that land use, so that one expects a gradual diminution of crowding as one moves outward. (The ~~central business district~~ CBD itself generally has a lower residential density than its environs. But in terms of the actual location of people — (at work, shopping, etc.) — it would probably have the highest density of all).

~~There are,~~ however, certain exceptions ~~which~~ are implied by the weight-falloff condition. Ideal weight depends not only on density, but on trip-taking propensities and the cost of moving the people involved. The weight of a retirement colony, for example, would be reduced for both these reasons: Retired people tend to be "light" because their incomes are low, and their earned incomes are very low; furthermore, their trips to the center of town will be relatively infrequent. Thus we would expect retirement colonies to locate farther out than most other land uses having their population density.

Again, consider the influence of family size on location of residence. Extra children generate more local trips - to schools, playgrounds, etc. - but probably few if any extra trips to the center of town. At the same time they increase the family's demand for floor space. Since local trips do not add to activity weight, the net result would seem to be that larger families tend to choose lower weight residential uses than smaller families. As a result, larger families should generally live farther from the center of town than smaller families. (This argument assumes that "other things" besides family size are equal - in particular, tastes and standard of living, as measured, say, by total income, or per-capita income, or something in between).

~~Let us~~ We now turn to the relation of family income and residential distance from the nucleus. The question of whether the rich or the poor live closer to the nucleus reduces, by the weight-falloff principle, to the question of whether the rich or the poor choose the heavier residential land use. Income will affect residential weight in two ways: via the density at which people live, and via the travel costs incurred per person.

The effect on density is well known. Low-income housing involves on the average more people per acre, for several reasons: There are fewer square feet of floor space per person; ~~a~~ greater proportion of housing is in the form of multiple-story dwellings; and there is less open space per person in the form of lawns, playgrounds, parking facilities, etc. (There are, to be sure, high-rise luxury apartments, but these are exceptional). The residential land uses of the poor thus tend to be heavier than those of the rich, hence closer to the nucleus.

The effect of income on travel cost is less clear. Travel cost resolves itself into the frequency with which one takes trips to the nucleus, and the cost per unit distance of a trip. Trip frequency in turn depends on steadiness of employment, tastes in recreation and shopping, etc.; the influence of income on these factors is unclear. ~~As discussed above, page~~ the cost of travel time rises with income, especially earned income. However ~~On the other hand~~, the rich are more likely to drive cars, ride taxis, etc., which compensate for these costs. While the overall effect is thus unclear, it may well be large, possibly large enough to overbalance the density effect. In this case the rich would live closer to the nucleus than the poor.[43]

This ambiguous conclusion is matched by ambiguous evidence. In developed countries the poor generally live closer to the center of town, but this is often reversed in the underdeveloped world and in past civilizations.[44]

The tendency for the poor to live closer to the center is abetted by the fact that cities grow outward, so that the older, more dilapidated, buildings tend to be located toward the center. The poor will gravitate toward the lower quality housing stock (see page above) and thus inadvertently settle close to the nucleus.[45] This factor, however, escapes the confines of the formal model of this section, because it involves a change in control of land over time: The original occupants of central-city housing hand it over to poorer successors as it deteriorates. The formal model, ~~on the other hand,~~ however, contemplates a single real-estate auction at "time zero".

In the above discussion, it has been difficult to avoid making statements of one-way causality among variables. This is misleading, since ~~one is~~ we are really dealing with the simultaneous determination of equilibrium values. For example, the housing choices of rich vs. poor families — (even assuming identical tastes) — will depend on the structure of rental-prices for the various qualities of housing, and these prices in turn depend on the actions of all land users in the system. For that matter, the level of family income itself is in part subject to choice, to be decided on jointly with the choice of housing quality, location, trip-frequencies, etc.

Similarly, take the relation of car ownership to location. An automobile acts as a general "levitator", so that one expects car owners to make lighter residential choices and therefore locate farther out. This makes car ownership appear to be the exogenous, causal variable. But one can also argue that families who decide to locate farther out find car ownership relatively advantageous, and are therefore more likely to buy cars. Actually, the problem for each family is one of choosing the most preferred configuration of several variables jointly, and it is in general not correct to argue that the choice of one variable causes the choice of another. We do find a positive correlation between car ownership and distance from the center of town, as expected, but no clear-cut causal relation has so far emerged, which again is to be expected in view of the above comments.[46]

~~Let us~~ We now turn from the structure of land uses to land values. As usual, $f(x,y)$ is the transport cost per acre incurred by a land use of weight y at distance x from the nucleus, and $p(x)$ is the land value per acre at distance x (area, weight and distance all being "ideal" quantities). We concluded above that if, for each y, the function $f(\cdot,y)$ is increasing and concave, ~~then~~ the function p will be decreasing and convex. Is this premise realistic, and, if so, is the conclusion borne out in real cities?

One difficulty that arises in trying to answer this question is that transport costs and land values are given in terms of physical, not ideal, acres and miles, so that a translation problem arises. For most urban land uses, one flat, welldrained parcel is about as suitable as another, and for these we ~~one~~ may identify real and ideal area. For the sake of argument we ~~let us~~ also identify real and ideal distances. (One way to judge the realism of this assumption is to see how well the isochrones, ~~-~~ the loci of points of equal travel time from the nucleus, ~~-~~ approximate to concentric circles). For fixed weight y, transport cost certainly increases with distance. Increasing congestion on ~~as one~~ approaches to the center of town tends to make it concave as well. That is, the closer one is to the nucleus, the more costly it is (in terms of time, mental strain, etc.) to travel a given physical distance; this makes f concave in distance. As for the conclusion, the most striking manifestations of the decreasing convexity of p are the extreme heights to which land values per acre rise in the central business districts of large cities.

These results must be interpreted with some care. "Land value" refers to time zero, where all land is assumed to be uniform (except for varying distances from the nucleus~~, of course~~). To be concrete, think of the original state as a vacant lot, with no capital improvements on it (except perhaps for drainage and leveling). Indeed, the formal model envisions just one omnibus real-estate auction at time zero, so that land values at other times are not even meaningful. In reality, of course, the real-estate market endures, with control of the same parcel passing from agent to agent. The question then arises: Must the cross-sectional distribution of land values at any future time conform to the Thünen pattern?

The answer is no. To take a not-unrealistic example, suppose someone invests heavily in a suburban plot, while an "inner city" plot is allowed to become "blighted". The former can easily become more valuable than the latter, reversing the original order. Only if one can isolate a "pure site value" might the original Thünen pattern be retained. (Note that the "improvement value" of a site can be negative, so that clearance of the site — which restores its original vacant condition — improves its market value).

Let us, finally, indicate some ways in which the assumptions of the Thünen model might be weakened in a realistic direction without invalidating its conclusions. We begin by introducing real-estate <u>taxation</u>. Suppose there is a <u>tax function</u> $t: S \times Q \to$ reals, where $t(s,q)$ is the tax liability incurred per acre by an agent running land use q at site s. More precisely, if agent n controls region S_n and chooses assignment ν_n over $S_n \times Q$, ~~then~~ his total tax liability is

15 ½ ⅙

$$\int_{S_n \times Q} t(s,q)\, \nu_n(ds,dq). \qquad (8.8.4)\ \cancel{(4)}$$

5a 6 This tax liability must be reflected in the preferences of the agents participating in the system. The natural way to do this is by a simple reinterpretation of preference condition (7.4); ~~(4) of section 7: Namely~~ viz., of two actions leading to the same allotment, the agent prefers the one incurring lower total costs, where costs include not only land costs, (7.3) ~~(3) of section 7,~~ and transport costs, (7.2) ~~(2) of section 7,~~ but <u>also</u> tax costs, (4) ~~above.~~

We ~~shall~~ do not attempt a fully rigorous discussion of the influence of this new factor, but instead follow the informal simplified approach of (7.72) ~~(72)~~ ~~to~~ to (7.75) ~~(75), section 7.~~ Let there be a social equilibrium, with agent n_i running land use q_i at site s_i, $i = 1,2$. The weight of q_i is y_i, and the distance of s_i from the nucleus is x_i, $i = 1,2$. Then

48 $$t(s_1,q_1) + \tilde{p}(s_1) + f(x_1,y_1) \leq t(s_2,q_1) + \tilde{p}(s_2) + f(x_2,y_1), \qquad (8.8.5)\ (5)$$

where $\tilde{p}$ is the land-value density function.

The argument for (5) is essentially the same as for (7.72) ~~(72) of section 7.~~ ~~Namely,~~ viz. agent n_1 has the option of switching one acre of land use q_1 from site s_1 to s_2. This leaves his allotment unchanged; hence, since he is initially in equilibrium, this switch cannot decrease his costs. Thus the left side of (5), which is the tax plus land plus transport cost incurred from running an acre of q_1 at s_1, does not exceed the right side of (5), which is the total cost he incurs from running q_1 at s_2.

Reversing the rôles of agents n_1 and n_2, we find that (5) remains true if the subscripts "1" and "2" are interchanged throughout.

We now make the assumption that the tax function <u>t</u> is <u>separated</u>. ~~That is,~~ i.e. we assume ~~there exist~~ functions

$g: S \to$ reals and $u: Q \to$ reals exist such that

$$t(s,q) = g(s) + u(q), \qquad \text{(8.8.6)} \quad \cancel{(6)}$$

for all $s \in S$, $q \in Q$. How realistic is this assumption? First, consider a Henry George-type tax on unimproved site value. This depends only on s, not on q, hence is of the form (6) (with u identically zero). Next, consider a tax on improvements only. It is not implausible that the improvement value of a given land use q will not vary much from site to site; in this case the tax function will be more-or-less independent of s, and again be in the form (6) (with g identically zero). (The principle of "treating equals equally" would tend to make such a tax independent of site in any case). Actual real-estate taxes are generally in the form of a sum of the two types just mentioned, hence again in the form (6).

There are a number of circumstances, however, in which (6) is not realistic. One example is when different land uses are taxed at different rates — say when industrial land uses are taxed more heavily than residential. For in this case the rate of taxation of site value depends on the land use occupying the site, and there is interaction between s and q. Another such case arises when Space is partitioned into several different political jurisdictions, each with its own tax structure. Thus, one jurisdiction may grant tax concessions to industry, while another may not; this leads to interaction between s and q. Even when the tax structures of different jurisdictions are similar and they differ only in tax levels, (6) is violated; for in this case s and q would combine multiplicatively rather than additively. Very roughly, then, (6) seems a good approximation within a single political jurisdiction, but not between them.

~~Let us~~ We now trace the implications of (6). Substituting in (5), we obtain

$$[g(s_1) + \tilde{p}(s_1)] + f(x_1,y_1) \leq [g(s_2) + \tilde{p}(s_2)] + f(x_2,y_1). \qquad \text{~~(7)~~ (8.8.7)}$$

Furthermore, if we first interchange subscripts in (5) and then substitute, we obtain (7) with its subscripts interchanged. Now (7) and its interchange differ from ~~(72)~~ (7.72) and ~~(73) of section 7,~~ (7.73) respectively, only in the fact that the function $\tilde{p}$ in (72)–(73) is replaced by $g + \tilde{p}$ in (7). The argument following ~~(73) of section~~ (7.73) 7 may now be repeated verbatim, except for replacing $\tilde{p}$ by $g + \tilde{p}$. We conclude, ~~then,~~ that the equilibrium assignment still satisfies the weight-falloff condition, even with the tax. Also, $g + \tilde{p}$ depends only on ideal distance h, so that ~~there~~ exists a function p:reals $\rightarrow$ reals satisfying

$$g(s) + \tilde{p}(s) = p(h(s)),$$

for all $s \in S$. The function p has the properties that, if $f(\cdot,y)$ is increasing for all y, then p is decreasing, and, if $f(\cdot,y)$ is concave for all y, then p is convex.

These results must be interpreted with care. The fact that weight-falloff is preserved does not mean that the tax has no behavioral effects. On the contrary, heavily taxed land uses (high u-values) will tend to occupy fewer acres, and may disappear entirely, to be replaced by lightly taxed uses. But the *order* in which land uses are ranged, by increasing distance from the nucleus, will not change. Heavyweight uses will still be closer than lightweight uses.

~~Note that~~ ~~it~~ it is no longer land values *per se* ~~which~~ that have the simple regularity properties of the Thünen system, but only land values *plus* pure site taxes. Thus, if one parcel is more heavily taxed than its neighbor, its equilibrium land value will be sufficiently below that of its neighbor so that the sum of value plus tax is about the same on both. This is essentially the classical conclusion that a pure site tax falls completely on the landlord, even if levied on the tenant. ~~(Actually,~~ this is a bit inaccurate. The tax has wealth redistribution effects ~~which~~ that reverberate throughout the system).

We conclude the discussion of taxes with an exercise. Show that if (6) is altered by adding the term $\theta(h(s),w(q))$ on the right, $\theta:\text{reals}^2 \to \text{reals}$ being any function having non-negative cross-differences, then all the conclusions above remain valid, with the single exception that, in determining the properties of p, the function $f + \theta$ should be used in place of f alone.

The remaining generalizations ~~we shall discuss~~ involve reinterpretations of the concepts "transport cost" and "activity weight" used in the formal model. The basic idea of the model, after all, is that people try to move their activities closer to the nucleus to economize on transport costs; "activity weight" measures the strength of this pull. Now suppose there were some other factor (not necessarily having anything to do with transportation), that ~~which~~ made it desirable to be closer to the nucleus. This factor would then operate as a "pseudo transport cost" and could be incorporated formally into the model provided -- and this is a strong assumption -- its effects could be summarized in the ideal distance function h already in use for transportation. The "weights" of the various activities would be adjusted upward to reflect this new attractive force, the size of the adjustment depending on the impact of the factor on the particular activity in question. Similarly, there could be a factor that ~~which~~ made it desirable to be further from the nucleus, and this would lead to a downward adjustment of activity weights. (One could even have mixed effects, some activities being attracted and others repelled from the nucleus because of some factor). We ~~shall briefly~~ discuss several examples of such factors.

Take danger, for instance. This comes in several forms. Danger from invasion and raiding parties has been important in the past, as the existence of walled cities testifies; it still is present today in "unpacified" countries. This danger is greatest at the periphery and diminishes as one moves toward the center of town. Hence it constitutes an attractive force and increases ideal weights. As for the differential effects, consider once again residential activities of rich vs. poor families. The rich will presumably be willing to pay more for a given increment of safety than ~~will~~ the poor. Hence the residential activities of the rich gain more weight than those of the poor, and this ~~accentuates the~~ creates a tendency for the rich to live closer to the center of town ~~than the poor~~.[47]

Under modern conditions this source of danger is minor. Two other sources, however, are quite important: danger from criminals and danger from thermonuclear attack. These have just the opposite pattern, being least dangerous at the periphery and most dangerous toward the center. (For a partial explanation of this, see ~~chapter 5~~, section 5.8). Hence they have a dispersive effect, reducing ideal weights. ~~Again,~~ the rich should be more sensitive to these influences than the poor, accentuating the tendency for the rich to live farther from the center than the poor.

Most non-transport factors seem to follow this latter pattern, adding a centrifugal rather than centripetal force, reducing rather than increasing ideal weights. This applies to other aspects of the "urban syndrome", such as pollution. The fact that pollution decreases as one moves outward makes it a dispersive force. The rich, being willing to pay more than the poor for a given physical decrease in pollution-intensity, will again tend to reside farther from the nucleus on this account.

A different kind of influence may be allowed for by relaxing the entrepôt assumption that all (non-local) trips must go through the nucleus. Suppose instead ~~that~~ there are also external trips, in which one travels to the outside world by heading directly away from the nucleus. Examples are furnished by certain types of outdoor recreation, such as pleasure driving in the countryside. These trips exert a pull opposed to that of the nucleus, hence ~~set~~ act as "levitators". The reduction in weight will be large for ~~those~~ people with a strong taste for outdoor recreation or other activities involving external trips, and ~~these people~~ they will tend to reside at the periphery.

As a final example, suppose some factor of production can be obtained on terms that vary systematically by distance from the nucleus. Specifically, suppose that wages decline as one moves away from the nucleus.[48] Then any land use with positive labor cost can reduce it by moving outward. The wage gradient therefore acts as a levitator. The land uses most affected will be those ~~which~~ that are most labor-intensive; to be precise, the reduction in weight among land uses is proportional to their labor/land ratios.

(A)

8.9. Thünen Systems and the Real World

The purpose of this section is, first ~~of all~~, to make a brief survey of some real-world Thünen systems, and, secondly, to discuss briefly the origin and function of such systems, and how they fit together in a hierarchy.

Whether a given real situation does indeed exemplify some theoretical concept is, in general, a matter of judgement. One rarely finds a pure case in which the assumptions of the theory are realized exactly: Instead there is generally more or less "noise" ~~which~~ that distorts the ideal conditions of the theory. If the "noise" is intense, the distortion may be so severe that the theory is useless for understanding what is going on. The question, then, is not whether the theory as interpreted is true or false, but whether it is a good or poor approximation to the situation in hand.

In the present case we ~~shall~~ consider a real-world pattern to be a Thünen system if it satisfies two conditions. First, there should be some indication of the concentric ring structure predicted by the weight-falloff condition. Second, there should be some indication that this structure arises via the central mechanism of the theory, ~~namely~~ viz., the attraction of the nucleus on the various land uses in operation.

It would be highly desirable to develop a series of indices of "goodness of fit" to ascertain in quantitative terms just how closely the following patterns <u>do</u> approximate to the ideal

Thünen pattern. To do this, however, would take us far off course. In the absence of such indices the following comments may be helpful. The approximation to the ideal should be best at intermediate distances from the nucleus, but poor when very close to the nucleus and at the periphery of the system.

There are several reasons why the fit should be poor at the periphery. The pull of the nucleus is very weak here, so that this systematic effect is more easily overshadowed by random "noise". Furthermore, the pattern tends to be disrupted by external forces – (such as the pull of nuclei of neighboring Thünen systems) – which are strongest at the periphery. Near the nucleus, on the other hand, the fit is poor because of a scale effect. The nucleus, after all, is not really a geometrical point as the model demands, but a region ~~which~~ that is small compared with the system as a whole (~~for example,~~ e.g., the inner central business district of a city). As distances become sufficiently small to be comparable in size to the nucleus itself, the direction of pull becomes uncertain and its strength attenuated, just as the pull of gravity weakens as one goes below the surface of the Earth.

The distinction between local and non-local movement is important here. The latter refers to trips to and from the nucleus while the former refers to short-distance trips ~~which~~ that do not affect the weight of land uses. Whether a trip is local ~~or not~~ depends on the particular Thünen system ~~one is~~ being investigat~~ing~~ed. Thus a trip to a neighborhood shopping center would be local in the context of the city as a whole, but non-local when the neighborhood itself is thought of as a miniature Thünen system. The distinction allows a great deal of flexibility in applying the apparently rigid requirements of the entrepôt model, that all trips must go to or come from the nucleus. The totality of local movements may far overshadow non-local movements; they undoubtedly do so for the larger Thünen systems. But non-local movements are focused on the nucleus, while local movements are diffuse in direction. This allows the former to exercise a systematic influence and lead to Thünen ring formation.

~~Let us~~ We now turn to real-world applications. (The following references are just a selection from a much larger number that could have been cited.) The bulk of our illustrations have been at the city level, with the ~~central business district~~ CBD playing the role of nucleus. The applicability of the Thünen model to cities was noted by ~~Walter~~ Isard (the germ of the idea can be traced back to Thünen himself).[49] Independently, the "concentric-circle" theory of city structure was promulgated by ~~Ernest~~ Burgess.[50] This was essentially an idealized description, stressing the (positive) correlation between distance from the center of town and socio-economic status of residents. No satisfactory explanation of this pattern was given, but, as we have seen, this relation can be incorporated into a modernized Thünen analysis.

We pass over the plethora of other studies of internal city structure — (by sociologists, geographers, economists and others) — and turn to the question of the existence of the nuclear attraction mechanism underlying the Thünen model. There are several possible approaches here. The most straightforward is to take a census of traffic flows. Many studies confirm the fact that a large fraction of all trips have the CBD as one terminus.[51] (The fraction appears to be declining, however, as suburbanization continues). A related approach examines the structure of the transportation grid, noting the extent to which it is "radial", focusing on the nucleus, vs. "peripheral", bypassing the nucleus. The predominance of a radial transportation grid is both a symptom of strong nuclear attraction — since transport arteries tend to get built along routes of heavy traffic flow — and a further influence strengthening that attraction. The same can be said about scheduled common-carrier routes. e.g. ~~For example,~~ bus routes in Queens, New York, run predominantly east-west, indicating that they serve as feeder lines for Manhattan commuter traffic.[52] Railroads in particular exhibit a strong radial pattern about large cities, extending like spokes of a wheel from the city-hub into the hinterland.[53] Furthermore, the city itself, especially if large, tends to be situated at a point of high natural "nodality", having good access to navigable rivers, mountain passes, the ocean, etc.[54] (An explanation for this will be given in ~~chapter 9,~~ section 9.4).

This tends to make regional traffic go through the city, making it a natural entrepôt, and underpinning the assumptions of the Thünen model.

From the city level ~~one~~ we can travel both upward and downward in scale. ~~Let us~~ We first go downward. The Thünen model may be applied to communities within a city. ~~E. Franklin~~ Frazier found that Harlem in New York City exhibited a ring structure that reproduced in miniature the Burgess concentric-zone model.[55] At this level, the attraction of the nucleus is likely to stem from the availability there of shopping and recreational facilities, and from the fact that the nucleus provides a transportation gateway between the community and the rest of the city.

Farm villages serve as nuclei for the surrounding fields. The nearer fields tend to be cultivated more intensively than the more distant, as befits a Thünen system. Here the attraction of the nucleus stems from its residences and associated local facilities; it may also serve as a collection point for farm products. An individual farm with uniform terrain will tend to have a similar structure, the nucleus being the farmhouse. The problem of laying out the farm efficiently, in fact, can be formulated as an allotment-assignment problem (if the farmer makes a few simplifying assumptions), and the solution to this satisfies the weight-falloff condition.[56]

The Thünen model applies roughly to theaters, arenas, and other facilities for public exhibitions. Here the nucleus is the center of attention — (stage, boxing ring, etc.) — and "transport cost" is incurred in the form of deteriorating quality of the view as one moves farther from the center. "Land values" are the prices of seats, and these tend to rise as one moves closer to the center. "Land uses" are persons occupying seats, and their "weights" are given by their willingness to pay for a given improvement in quality of view. The "heavy" people will then gravitate to ringside, while the "light" people become groundlings.

On ~~one's~~ a desk or workbench, a rational layout would find the more frequently used items closer to hand. One can give many more examples of such microscopic Thünen systems, called into existence by nuclear attraction. ~~Let us~~ we now, however, drop these small-scale systems and direct attention upward from the city level.

First ~~of all~~, the influence of a city generally extends far beyond its political boundaries. This influence is reflected in the correlations that exist between distance from the city and such variables as population density, land values, intensity of farming, income, etc. This "metropolitan" system is merely the continuation outward of the city system; the latter — (to which we have been devoting most attention) — constitutes the inner rings of the complete system. When taking the "metropolitan" point of view, it is often convenient to think of the entire city as being the nucleus. This is the approach of Thünen himself and is implicit in some of the citations given above.

Gal 32-1

Above the city-metropolitan level lies what might be called the sub-continental level. The system may be mostly contained in one country, as in the United States, or may embrace several countries, as in Western Europe. The nucleus is the "industrial heartland" of the region, which has the greatest concentration of population and capital goods and provides the major market of the system. In the United States this is the "Northeastern industrial belt".[57] In Europe it is the Rhine-Ruhr complex. Land values, population density, and the "weight" of land uses all tend to decline as one moves away from the heartland. Also, per-capita incomes tend to decline. This last fact establishes a link between sub-continental Thünen systems and the tendency for developing economics to develop a "dual" structure, with a high-income sector utilizing modern technology and a lagging low-income sector using traditional methods. Very roughly, the "modern" sector will occupy the inner rings of the system, while the "traditional" sector occupies the periphery. For example, a United Nations study points up a very general tendency for the low-income sector within each Western European country to be that part ~~of it~~ farthest from the Rhine-Ruhr complex.[58]

One very suggestive aid for delineating Thünen systems at this macro-level is John Q. Stewart's concept of "potential" (not to be confused with the potentials in the measure-theoretic transportation problem of section 7.5). For the most general definition, start with a measure space (S, Σ, μ) on which is defined a metric, $h: S \times S \to$ reals, $h(s, \cdot)$ being measurable for each $s \in S$. Then the (Stewart) potential for μ and h is the function p with domain S given by

$$p(s) = \int_S \frac{\mu(dy)}{h(s,y)^k}, \qquad \cancel{(1)}\ (8.9.1)$$

where $k = 1$. S is usually a portion of the surface of the Earth, and h is usually Euclidean distance (perhaps great-circle distance if the curvature of the Earth over S is significant). μ is usually population or income measure (in which case p is called population or income potential, respectively), but other measures, such as employment or retail sales, are also used.

The potential p is, in effect, a spatial moving average of measure μ, and may be thought of as an index of "generalized closeness" to μ. In using p to delineate Thünen systems, one takes p to index "closeness" to the nucleus, so that the iso-potential contour lines are taken to be loci of equal ideal distance. The justification for doing this is empirical: Potential has a high correlation with many variables one expects to be related to ideal distance in a Thünen system, and the contour maps of potential look reasonable in terms of where the Thünen rings *should* lie. Unfortunately, no good theoretical reason has been given as to why a function of the form (1) should have these properties.

(There is a very crude argument for setting the exponent k in (1) equal to 1, as is almost always done. Setting k equal to 0 obviously makes p a constant, a clear case of over- smoothing. ~~On the other hand~~ But (assuming S is the Earth's surface, and that μ has a continuous density function with respect to Euclidean area) ~~one~~ we can show that, as k approaches 2 from below, p becomes proportional to the density function of μ; thus ~~one is~~ we are, in effect, reproducing μ, a clear case of undersmoothing. The reasonable values for k thus lie between 0 and 2, and 1 seems a good compromise. This argument is weak, and it would be interesting to determine whether modifications of (1) give better fits).

Taking S to be the United States (excluding Alaska and Hawaii) one finds that the maps of income and population potential are similar. The national peak is in the New York City area, and the contours are roughly concentric about this point clear out to the Rocky Mountains. West of the Rockies, a much lower secondary system has emerged, centered on California. The primary system has persisted in main outline for more than

59 a century.[59] This may be taken as an estimate of the sub-continental Thünen structure for the United States.

At the highest level of all, ~~one~~ we ~~treats~~ treat the entire world as a Thünen system. Here the nucleus is the "North Atlantic Heartland" of Northeastern United States and Western Europe. One way of assessing the structure of the system is by examining international trade and travel, the bulk of which originates or terminates in this region. (At the world level, it is reasonable to assume that most "non-local" trips cross national borders, and are picked up in the international transaction accounts). Another approach uses potentials. ~~William~~ Warntz has constructed a world income-potential map (using great-circle distances).[60] This shows twin peaks, on the two sides of the North Atlantic, the world peak being in the New York City area. The contours are, for the most part, roughly concentric about the Heartland, with a much lower secondary system beginning to emerge, centered on Japan.

This picture of a world Thünen system evokes some familiar echoes. The period of high colonialism (say 1880-1914) has often been thought of in similar terms, especially by Marxian analysts. Here the mother countries of Western Europe struggle for colonies that ~~which~~ function as sources of raw materials, and as outlets for manufactured goods and investments; the mother countries are the centers of managerial and financial control. As a result, trade becomes strongly polarized into the radial Thünen pattern.

~~Of course,~~ this system transcends the Thünen model we have been discussing, since it involves military and political factors operating outside the market system. Nonetheless, there is reason to believe that a similar pattern would have evolved even under a single world free-market system. "Colonial" relations did arise in the U.S. sub-continental Thünen system, between Eastern merchants and "Wall Street" on the one hand, and Western farmers and Southern planters on the other. And on the metropolitan level one finds similar relations between city banks and hinterland farmers mortgaged to them.

In this same vein ~~one~~ we recalls Marshall Lin Piao's analogy in which the Western world corresponds to the "cities" and the underdeveloped world to the "countryside".[61] In effect, he is comparing the world Thünen system with the Chinese sub-continental Thünen system, with the thought that the laws (of political struggle) ~~which~~ that operate on one level should operate on another.

Consider the concept of <u>ideal distance</u> in the context of the world Thünen system. We have argued ~~above~~ (sections 2 and 3) that economic distance, based on unit transport cost, is the proper concept to use in Thünen models; this led us to expect certain characteristic distortions in the Thünen rings, such as elongations along major transport arteries and isolated pieces around points of access to the transportation grid. These same arguments apply to Thünen systems at any level. 8.2 and 8.3

At the world level, ~~there~~ still remain considerable differences between ideal and physical (i.e., great-circle) distances, although the advent of air and automotive transportation, as well as radio communication, has tended to make ideal distances conform more closely to physical distances than used to be the case. The ideal/physical distance ratio tends to be relatively low across bodies of water, and to be relatively high across mountains, deserts, and regions characterized by very cold climates, endemic diseases, and unsettled political conditions (pirates, bandits, guerrillas, ~~etc.~~). Ideal distance takes a discontinuous jump across national boundaries, because of tariff and customs barriers, exchange controls, travel restrictions, etc.

~~Of these,~~ let us concentrate on the implications of land vs. water travel. The advantage of water was very great in the pre-railroad ~~19th~~ nineteenth Century, when the shortest route from New York to California lay around Cape Horn. As a result the oceans of the world played a role in the world Thünen system similar to that which high-speed radial highways play in a city-metropolitan system. One consequence was that the interiors of continents tended to be farther from the North Atlantic Heartland, in ideal distance, than their coastal regions. The coast of Australia was closer to London than the interior of Africa was.

~~One~~ We would then expect the interiors of continents to have lighter land uses than their coastal regions, by the weight-falloff condition. A rough index of land-use weight is population density. And, indeed, an estimated 2/3 of the human population live within 300 miles of the sea; 56% of the population live at altitudes below 200 meters, on just 28% of the

62 world's land area.[62] (~~There are~~ other factors besides mere distance from the sea ~~which~~ keep continental interiors relatively empty, however, e.g., dryness, rough terrain, and cold).

583 Furthermore, continents are internally differentiated in terms of sea-access. In particular, land along navigable rivers will be relatively close to the world-nucleus in ideal terms, and should carry heavier uses than less accessible land;

a sort of continental "urban sprawl" functionally similar to that ~~which~~ spread~~s~~ing along intercity highways.

The various potentials (1) should presumably be calculated in terms of ideal distances h rather than physical distances, since this will yield a more accurate index of "generalized accessibility." Using ideal distances would strengthen the impression that we are dealing with a single world center rather than an American-West European bipole, since the North Atlantic Ocean gap would shrink.

This comple~~x~~tes our ~~brief~~ survey of real-world Thünen systems. ~~Needless to say,~~ the discussion has been impressionistic, intended to establish a framework around which more substantial quantitative work can be organized.

#

18 We now turn to the discussion of Thünen systems over ~~t~~Time rather than over Space. We describe a number of idealized historical patterns, into one or more of which most real-world Thünen systems should fit, approximately.

Our basic model does have dynamic features, since the concept of "land use" allows for non-stationary import-export patterns over Time. ~~Indeed,~~ certain dynamic phenomena — such as land speculation and suburbanization — have already been discussed and partially explained in terms of the model. Nonetheless, many aspects of the development of real-world Thünen systems fit into the model poorly or not at all. These include transportation construction, the origin and changing functions of the nucleus, and population movements and the consequent transfer of control over land.

Actually, many aspects of population re-distribution can be incorporated into our model without undue strain. Births, deaths, aging and other changes of state are accounted for in the capital structure of the various land uses in operation. (We have been concentrating almost exclusively on the export-import component of land uses in this chapter, and ignoring the capital component, because the latter makes no direct contribution to land-use weight.) Commuting has ~~already~~ been discussed at length. This leaves migration ~~to be discussed~~.

First, we make a point of definition. There is no sharp distinction between migration and commuting. The ultimate data ~~one is dealing~~ we with here are contained in a person's itinerary, the function giving his location at each moment of his existence. If this function is roughly periodic over a certain interval, ~~one~~ we speaks of "commuting". If it changes more-or-less permanently, ~~one~~ we speaks of "migration". There are various intermediate cases.[63]

In any case, both migration and commuting are accounted for by examining a person's trips. We distinguish, as above, between local and non-local trips, local trips being relatively short distance and not touching the nucleus, non-local trips being those going to, from, or through the nucleus. ~~As already discussed,~~ the distinction depends on the system of reference. Thus, a rural-to-urban migration trip would be considered non-local in the context of the destination city Thünen system, but local in the context of the world Thünen system. Non-local trips contribute to the weight of the land use at which they originate or terminate; local trips do not. But even for non-local trips, migration — as the term is commonly understood — probably makes just a minor contribution to weight; for it is non-repetitive, and the cumulative impact of commuting trips and goods shipments will tend to swamp it.

The significance of migration for the evolution of Thünen systems lies, rather, in its influence on the capital-structure of land uses. (This influence is, of course, the net result of migration, births, deaths, aging, etc. Construction and mining — including improvements, maintenance, scrappage, and demolitions — play a similar rôle in the capital-structure of non-human resources). When a land use enters into a more intensive phase of exporting and importing, this will in general be accompanied by a rise in resource density on the site: more people, plant, equipment, inventories per acre. We may expect net in-migration at the beginning of this phase. Similarly, a phase of reduced intensity should be accompanied by net out-migration.

Consider suburbanization, for example. This was explained in terms of the tendency for lighter land uses to have longer initial periods of vacancy. By the weight-falloff condition, this implies a general tendency for land more distant from the nucleus to remain vacant longer than land close in. Inmigration should occur about the time when this initial idle period comes to an end. Thus we may expect a ring of intensive inmigration, the ring itself expanding away from the nucleus over time. The migrants themselves may originate from points closer to the nucleus, from points farther out, or as "immigrants" from outside the Thünen system.

The predominant direction of flow of migration provides an important principle for classifying Thünen systems. *Condensation* systems are those in which the main flow of migration is inward, toward the nucleus. *Dispersion* systems are characterized by outward migration, away from the nucleus. One rarely deals with a "pure" case; e.g. ~~For example,~~ at the city level, suburbanization is mainly a dispersion process, the old urbanites moving out; but at the same time, the condensation process of rural to urban migration still goes on. Furthermore, the same system might be dispersive during one epoch of its history and condensive during another. Nonetheless, the distinction is useful, and we now discuss the conditions that can be expected to yield one or the other of these processes.

A dispersion system is likely to arise when the surrounding countryside is vacant, or occupied by a sparse aboriginal population living at a low technological level. The United States subcontinental Thünen system furnishes one example; this was peopled by "westward expansion", which is also "outward dispersion" from the eastern nucleus. A second example is the eastward expansion of Russia into Siberia (this may perhaps be thought of as dispersion into the eastern periphery of the western European subcontinental system.). In the U.S.

case, an additional force making for dispersion was the great tide of migration from Europe ~~which~~ that entered at the nucleus (so that, in a sense, this was dispersion into the western periphery of the Western European system).[64]

A condensation system is likely to arise, ~~on the other hand,~~ however, when the surrounding countryside is occupied by a dense initial population. (Incidentally, this assumption does not invalidate the model used in this chapter. It is required only that initial conditions be uniform over Space, not necessarily that there be vacancy everywhere.) The major type is rural-to-urban migration. Condensation also characterizes "dual" economies, with people in the traditional sector migrating into the small modern sector.

The spatial distribution of income will, in general, differ in these two kinds of Thünen system. There is a general tendency for migrants to move from regions of low per capita incomes to regions of high per-capita incomes. Hence a condensation system should be characterized by higher incomes toward the center, and a dispersion system by higher incomes toward the periphery. These expectations -- (especially the latter) -- are by no means certain. For migration is selective, and it is entirely possible that the migrants improve their position even if average income (including income of non-migrants) is lower at destination than at origin. In particular, the "pioneers" who migrate in a dispersion system must do without the amenities of civilization, and tend to be drawn from the lower income strata.[65] Thus even a dispersion system might have declining income with increasing distance from the nucleus.

66 These two kinds of Thünen systems appear to be exemplified in two well-known economic development models. The model of W. Arthur Lewis[66] leads to a condensation system. Here labor in the traditional sector of an underdeveloped economy is available in completely elastic supply to the modern sector, which creates a ceiling on wages in the latter sector. Expansion of the modern sector pulls in surplus labor from the countryside. The model of Frederick Jackson Turner,[67] on the other hand, leads to a dispersion system. This is the mirror-image of the Lewis model, for Turner postulates an unlimited supply of land at the frontier of settlement; this creates a floor ("safety-valve") on wages at the center, for any tendency of wages to fall will lead to out-migration to the periphery.

start 32-5 ~~Note, by the way, that~~ neither condensation nor dispersion requires long-distance migration. The same population redistribution can be accomplished by a series of short-distance moves (short in comparison to the radius of the Thünen system). This "step-by-step" movement is in fact very common. People move, leaving a vacancy into which people behind them eventually move, who in turn leave a vacancy, etc. (Example: the "filtering" process by which housing gets handed down from higher to lower income families, accompanying movement into the suburbs). Or, instead of a series of "pulls" there may be "pushes": People move, which leads to crowding and conflict at their destination, and the driving out of people ahead of them; these in turn push out others, etc. (The European Völkerwanderung was partly of this type).

~~Let us~~ We now turn ~~briefly~~ to the question of origins of Thünen systems. In the condensation case, there appear to be (at least) two ways in which these systems originate. One is the passing of a certain <u>threshold</u> of concentration. This is a compound of density of population, size of per-capita income, and mutual accessibility. (The peak regional income potential provides a rough index of concentration.) At the threshold, the market is just wide enough to make certain centrally-located enterprises economically viable. These come into existence and start a snowballing process, involving the founding of linked enterprises; deepening division of labor; construction of plant, utilities, transportation arteries, and communication links; in-movement of population; innovations; etc.[68] The second way is for some autonomous localized innovation to occur, such as the discovery of a mineral deposit or the opening of trade relations with the outside world. The center itself tends to be founded at a point of high general accessibility, either by virtue of natural "nodality" (river confluence, good harbor, etc.) or because of previous transportation construction. And the very act of founding the center leads to further radial transportation construction ~~which~~ that artificially increases the advantages of the site.

Dispersion systems typically originate through an incursion from the outside world ~~—~~ (say first by explorers, missionaries or soldiers, then traders, then settlers). The nucleus is then ~~apt~~ likely to be near the original port of entry ~~into~~ to the region.

Once the snowballing process begins, the nucleus exerts an attractive force on the entire region, which tends to pull land uses into the Thünen ring pattern. The nucleus acts as trade center, gateway to the outer world, source of specialized goods and services, and major employment center.

As the system matures, a number of forces arise to retard growth at the center, perhaps even to halt and reverse it. Vacant land becomes scarce; traffic congestion grows. The capital plant becomes aged and obsolescent; pollution rises with density of population. As the world develops, the original location of the center may become less advantageous; mineral deposits are played out, harbors become silted. At the same time, concentration in various parts of the hinterland reaches a threshold of its own, and new competing centers arise. (These are ~~apt~~ likely to be located near points of access to the transportation grid feeding the original center, ~~such as~~ e.g., railway stations, highway interchanges, and airports).

18 These processes go on simultaneously in Thünen systems of all levels. ~~Let us~~ We now change our point of view, and study the interrelations among these systems. The following idealized model appears to be a fair approximation to many real-world

situations. Thünen systems form a hierarchy, in the sense that, for each system there exists a higher-level system to which it stands in a certain subordinate relation. (The exception, ~~of course~~, is the world Thünen system, which stands at the apex of the hierarchy). Let the system occupying region S_1, with nucleus s_1, be subordinate to the system occupying region S_2, with nucleus s_2. This means, first ~~of all~~, that S_1 is a sub-region of S_2; secondly, trips between points of S_1 and nucleus s_2 go through nucleus s_1. That is, s_1 functions as a gateway between its own hinterland and the superior nucleus s_2.

32-69 This "gateway" function comes in several diverse forms. Nucleus s_1 may be the site of a major railway station or highway intake point, so that long-range traffic funnels through it. Out-of-region telephone calls may go through a local exchange at s_1. Wholesalers or manufacturers at s_2 may ship to retailers at s_1, who in turn sell to customers in the hinterland S_1. Conversely, s_1 may serve as an assembly point for goods produced in S_1, these goods then being shipped in bulk to s_2. Mail addressed to points in S_1 may be shipped from s_2 to a local post office at s_1, and thence distributed by letter carriers; outgoing mail follows the opposite route. Businesses and government agencies may have regional offices at s_2 and local offices at s_1; complaints, information, orders, merchandise, tax forms, etc., which flow between hinterland residents and these organizations may then be channeled through these local offices.

Now let the superior Thünen system (S_2, s_2) itself be subordinate to the higher level system (S_3, s_3). Trips, shipments, and messages between points in the hinterland S_1 and nucleus s_3 will then go through both intermediate centers, s_1 and s_3. The picture that emerges is that of a "chain of command", in which messages from one unit to another go first up the chain to the lowest unit superior to both of them, then down the chain to the receiving unit. Approximations to this structure are found in the circulation of mail through the postal system, of messages through the telephone system, ~~of~~ and checks through the bank clearing system, as well as in the administrative structure of complex organizations.

Trips between two sites in S_1 are to be considered local when analyzing the system (S_2, s_2), because such trips are short-circuited through the nucleus s_1, and nucleus s_2 exerts no extra pull on the land uses generating them. Similarly, for the system (S_3, s_3), the much wider collection of trips between sites in S_2 are to be considered local. The same land use thus gets lighter and lighter as it is referred to higher and higher level systems.

The general conclusions of Thünen analysis — (the weight-falloff condition, the declining convex structure of land values, etc.) — apply at each level of the hierarchy. As one goes up the chain of systems, a broader picture emerges, but more detail is lost, because more trips are relegated to the "local" category and ignored. Thus an analysis at several levels is needed to get the full picture.

What cannot be explained by Thünen analysis alone is the hierarchical structure itself. We ~~just~~ list a few of the factors involved. There are, first ~~of all~~, bulk economies in transportation and communication. Thus one builds a small number of channels ~~which are~~ used collectively, rather than a separate channel between every possible pair of sites. This creates artificial nodes around which centers can grow. A distributional apparatus arises for collecting, storing, assembling, shipping in bulk, and disassembling into retail lots. Secondly, various enterprises exhibit scale economies which, when combined with distributional costs, lead to a diversity of preferred spacings for different industries. (This will be discussed further in ~~chapter 9,~~ section 9.6).

The preceding discussion of hierarchy touches central-place theory at many points.[69] ~~Indeed,~~ the central-place and Thünen approaches are complementary, in that the latter concentrates on the structure of land uses in the field, while the former concentrates on the activities at the various nuclei. One major theme in central-place theory is that ~~there exists~~ a total ordering (~~with~~ allowing indifference) of both activities and centers exists, such that nth-level activities are present only in nth and higher level centers. We make ~~just~~ two ~~hurried~~ brief comments about the relations between these approaches. First, while the Thünen analysis is a social equilibrium, based on the interactions of agents choosing most preferred actions, the central-

place analysis does not rest on this foundation. Instead, it is an idealized description, and it is not at all clear, for the most part, how the patterns it postulates arise from the interactions of rational agents. Second, one of the few exceptions to this last generalization arises from the Thünen analysis itself. We have seen ~~(page above)~~ that with a limited-access transportation grid, the weight-falloff condition leads to the ~~above-mentioned~~ "central-place" pattern of
32-1 activities and centers. (The centers form at the access points; the higher order centers are those closer to the nucleus in ideal distance; the higher order activities are the heavier ones.) While this model alone is inadequate, it suggests that a modified Thünen framework, incorporating such hierarchy-producing assumptions as bulk economies in transportation, might prove to be an adequate theoretical underpinning for the central-place model.

Finally, a word about the dynamics of the Thünen hierarchy. One index of the "importance" of a particular Thünen level is the ratio of non-local to local trips, for this measures the degree to which the nucleus participates in the overall functioning of the system. The data for this index are generally not available, but one can often make do with a cruder index ~~which is~~ presumably correlated with the original. For example, the "importance" of the world Thünen system can be estimated by the ratio of the value of total international trade to gross ~~total~~

70 world ~~income.~~ product. This ratio rose ~~greatly in the Nineteenth Century.~~ from, say, 0.03 in 1800 to 0.33 in 1913 (but declined somewhat thereafter).[70]

no ¶ This illustrates one very long-term trend: the rise in importance of the higher level centers at the expense of the lower level. This trend has certainly not been steady, but if ~~one~~ we ~~examines~~ examine human history in terms of millenia rather

71 than centuries, it is palpable enough.[71] It is associated with other trends ~~which~~ that have been much commented on, such as the deepening division of labor, the rise of national states and incipient world organizations, and the shift from agricultural to blue collar to white collar work, especially of the information-processing variety. One basic causal factor in all this is doubtless the reduction of the real cost of transportation and communication, especially for long distances, but the

72 detailed causal interrelations are still obscure.[72]

FOOTNOTES - CHAPTER 8

773 1. [1]Von Thünen's Isolated State, P. Hall, ~~editor~~, trans. C. M. Wartenberg, ~~translator~~ (Pergamon Press, New York, 1966). The German original was first published in Hamburg, 1826.

---

778 2. [2]These phenomena are sometimes referred to as the "shrinking globe". We find it more convenient to leave the globe unshrunk and to represent them by a reduction in the other cost component, ideal weight. Thus one might speak of "levitating resources" as resources in general get "lighter" over time.

---

780 3. [3]The notions of "ideal distance" and "ideal weight" (the latter not depending on Time) stem from Alfred Weber (1909). While the "ideal weight" concept has been widely accepted, location theorists have severely criticized "ideal distances". See E. M. Hoover, Location Theory and the Shoe and Leather Industries (Harvard Univer~~sity~~ Press, Cambridge, 1937), ~~page~~ 40 ~~note~~ 10; W. Isard, Location and Space-Economy (MIT Press, Cambridge, 1956), ~~page~~ 109; ~~and especially~~ T. Palander, Beiträge zur Standortstheorie (Almqvist och Wiksell, Uppsala, 1935), pp. ~~pages~~ 195-199. This is curious. It is true that "ideal distances" will generally be non-Euclidean, so that special constructions based on the Euclidean metric cannot be used. But it should be clear that the "ideal" concepts of weight and distance are entirely on a par, and in fact only defined conjointly.

---

[4]We ~~are~~ tak~~ing~~ certain notational liberties in re-arranging the order of the component spaces. No confusion should result from this, since there are no repetitions among these components.

---

[5]An "allotment" constraint will be added later, but this constraint is "artificial" in a sense to be discussed below.

---

[6]If we are investigating the neighborhood Thünen system centered on the cluster of local facilities, rather than the structure of the city as a whole, ~~then~~ these local movements should be counted.

---

[7]If the roof is utilized, there are N + 1 surfaces, the topmost being unsheltered. "Open space" may be thought of as a zero-story structure whose "roof" is the surface of the Earth.

---

[8]This refers to "voluntary" travel. For "involuntary" travel — as by children, prisoners, and military personnel — the relevant valuation is not done by the traveler himself.

---

[9]This self-assessment feature of transportation cost raises certain questions concerning the welfare implications of the entrepôt model. We ~~shall~~ comment on this later.

---

[10]G. S. Becker, "A Theory of the Allocation of Time", Economic Journal, 75:493-517, September, 1965.

---

805 ¶11. [11] W. Alonso, *Location and Land Use* (Harvard University Press, Cambridge, 1964), ~~chapter~~ 1 and ~~pages~~ pp. 101-105. The term "friction of space" dates from R. M. Haig, 1926; it is sometimes taken to include total land value as well as total transportation cost.

gally 32-8

807 ¶12. [12] Note that the *f* appearing in (4), whose domain is two-space, does not have the same form as the unit transport cost function *f* of chapter 7, whose domain is A × B. The analogue of the latter is the composite function *f*(*h*(·), *w*(·)), with domain *S* × *Q*. However, below we ~~shall~~ convert this problem into another for which the *f*'s *do* correspond.

807 ¶13. [13] B. H. Stevens, "Location Theory and Programming Models: The Von Thünen Case", *Regional Science Association Papers*, 21:19-34, (1968). The model is on ~~page~~ 26. The objective here is to maximize total bid rent or profit, rather than to minimize total transport cost, but the formal structure is the same. M. Beckmann and T. Marschak, "An Activity Analysis Approach to Location Theory", *Kyklos*, 8:125-141, (1955), seem to have envisaged this model in a remark on ~~page~~ 128.

809 ¶14 [14] Agricultural programs often contain acreage allotment restrictions, but these typically apply farm-by-farm, not on a global basis for all Space as in (6).

[15] Such as: E. S. Dunn, Jr., The Location of Agricultural Production (~~University of~~ Florida Press, Gainesville, 1954); R. F. Muth, "The Spatial Structure of the Housing Market," ~~Regional Science Association Papers~~, 7:207-220, (1961); L. Wingo, Jr., Transportation and Urban Land (Resources for the Future, Inc., Washington, D.C., 1961); ~~W.~~ Alonso, Location and Land Use ~~(Harvard University Press, Cambridge, 1964)~~; E. S. Mills, Studies in the Structure of the Urban Economy (Johns Hopkins Press, Baltimore, 1972), ~~chapters~~ 5 ~~through~~ 8, ~~chapter 4 of Mills~~ is a good survey of other models.

---

[16] Measures on the plane satisfying the northwest (or rather, southwest) corner condition have been studied from another point of view by M. Fréchet. See W. Feller, An Introduction to Probability Theory and Its Applications, vol. ~~II~~ 2 (Wiley, New York, 1966), ~~pages~~ pp. 162-163, problem 6.

---

[17] We write the potential as the pair (p,k), rather than ~~as~~ (p,q) as in chapter 7, to avoid confusion with points $q \in Q$. The stated definition is for variant I of the transportation problem, which we are using. Other variants add non-negativity conditions on p or k. Also this is a "wide-sense" definition, so that there is no requirement that the integrals $\int_S p\, d\alpha$, $\int_Q k\, d\beta$ be well-defined and finite. The same remarks apply to the next definition.

~~15a For the product function f(x,y) = xy we need not assume half-boundedness; for G_i may be chosen to lie in a rectangle $E_i \times F_i$ with h bounded on E_i and w bounded on F_i, i = 1,2, making all four integrals in (14) finite.~~

poh | 849 ¶ 18. [18]Here $p \circ h$ is the composite function whose value at $s \in S$ is $p(h(s))$; Similarly for $k \circ w$.

851 ¶ 19. [19]When X and Y are constructed in this manner, the ranges of h and w may not be respectively contained in them. We shall deal with this minor complication in due course. For the time being ~~let us~~ we concentrate attention exclusively on the plane, and forget about the original allotment-assignment problem.

856 ¶ 20. [20]We need not distinguish between open, closed, etc., intervals here, since the value of a function at isolated points does not affect its measurability.

874 ¶ 21. [21]The plane in which the rectangle of support lies should not be confused with the plane in which S lies. The former is an intrinsic feature of any allotment-assignment problem; the latter is an accidental feature of the special allotment-assignment problem now under discussion.

874 ¶ 22. [22]This ambiguity of the additive constant is characteristic for variant I transportation problems, of which allotment-assignment is a special case. It has an economic interpretation, as we shall see.

875 ¶ 23. [23]If the allotment-assignment problem were formulated with an areal capacity inequality constraint (variant II or IV rather than I), this constant would be zero: Unused land is valueless.

878 ¶ 24. [24]The New York Metropolitan Region is fragmented by bodies of water. Yet population density maps indicate fairly close conformity to a Thünen ring pattern (centered on Midtown Manhattan) ~~which~~ that is unaffected by these interruptions. This is just what one would expect from the above analysis.

(863) ¶25. [25]A more careful discussion would note that the constructed allotment-assignment problem introduces extraneous solutions. ~~For example,~~ e.g. a 50-50 mixture of densities 10 and 30 is not the same as a density of 20. (Mixtures are meaningless for the allocation problem.) Hence the last theorem is stronger than it looks, since ν^{o} is optimal not only against assignments ν of type (17), but against "mixtures" as well.

902 ¶26. [26]We have implicitly assumed that all ~~functions and measures~~ integrals discussed are ~~bounded,~~ finite, to avoid pseudomeasure complications.

903 ¶27. [27]A borderline is a locus of constant h-value. These are "isodapanes" or "isovectures" of the Weber problem.

904 ¶28. [28]Also, under these conditions, g has positive cross-differences iff (32) holds, and the inequality is strict on a set of points (x,y) ~~which~~ that is dense in the plane. (To prove these statements, use the mean-value theorem). Incidentally, the existence of $D_2[D_1g]$ does not imply the differentiability, or even the continuity, of $g(x,\cdot)$, for any x.

907 ¶29. [29]P. A. Samuelson, Foundations of Economic Analysis (Harvard Univ~~ersity~~ Press, Cambridge, 1947), ~~page~~ p. 183.

[30] g_n in (7) is unique only up to a region of ν'-measure zero. For (8) to be well-defined, it must give the same value for any function g_n satisfying (7). It turns out that equilibrium μ must be absolutely continuous with respect to ν', which guarantees uniqueness in (8).

---

[31] Recall that ordinary summation of measures, as in (6), is defined iff all summands are over the same measurable space, while direct summation, as in (9), is defined iff the universe sets of the summands are disjoint from each other.

~~[31] For notational convenience we replace boldface Q by ordinary Q in the following formal discussion.~~

---

[32] ~~The~~ condition (4) on preference orders must be correspondingly modified, since "cost" may no longer be representable as a real number. The natural generalization is to interpret "$c_1 \leq c_2$" as referring to standard ordering of pseudomeasures. Cf. the real-estate market of chapter 6. We omit detailed discussion of this point.

---

[33] These difficulties motivate the "measure space of agents" approach, the number of agents being now uncountable, and each literally having zero influence. But this creates new conceptual difficulties of its own. See Section 6.7.

---

953 ¶ 34. [34] The following development is essentially that of ~~my~~ *Essays in Spatial Economics*, ~~pages~~ pp. 170–188. Ph.D. diss., Columbia University, 1967.

960 ¶ 35. [35] ~~W.~~ Alonso, *Location and Land Use* ~~(Harvard University Press, Cambridge, 1964)~~, ~~pages~~ pp. 101–105. ~~cf. page~~ above. See note 110

961 ¶ 36. [36] See, ~~for example~~ e.g., D. M. Winch, *The Economics of Highway Planning* (Univ~~ersity of~~ Toronto Press, Toronto, 1963).

962 ¶ 37. [37] ~~Incidentally,~~ the fact that total transport cost as given by ~~(2) of section 7~~ (7.2) is a *linear* function of assignment ν, hence of resource flows, indicates the absence of external effects. Thus the introduction of external effects would vitiate not only the ethical conclusions from the model, but the model itself.

962 ¶ 38. [38] B. J. L. Berry and A. Pred, *Central Place Studies*, 2nd print. (Regi~~onal~~ Sci~~ence~~ Resear~~ch~~ Instit~~ute~~, Philadelphia, ~~second printing~~, 1965). (See pages 000– ~~above~~.

962 ¶ 39. [39] To answer these questions, one needs the more elaborate models of Mills, Wingo, Alonso, Muth, Dunn, and others. See ~~page , foot~~note 15 ~~above~~.

964 ¶ 40. [40] R. Turvey, *The Economics of Real Property* (~~Geo.~~ Allen & Unwin, London, 1957), ~~pages~~ pp. 16–18. The "Sheridan-Karkow formula" for the rental value of office space shows rent increasing with height; see *Buildings*, December 1959, ~~page~~ 30.

966 ¶ 41. [41]cf. R. Sinclair, "Von Thünen and Urban Sprawl", Annals ~~of the~~ Association of American Geographers, 57:72–87, ~~March, 1967,~~ (1967); and E. M. Hoover, The Location of Economic Activity (McGraw-Hill, New York, 1948), ~~pages~~ pp. 170–171, for examples. An alternative explanation of leapfrogging is based on the presence of imperfect information in the real-estate market: see W. R. Thompson, A Preface to Urban Economics (Johns Hopkins Press, Baltimore, 1965), ~~pages~~ pp. 326–327.

---

966 ¶ 42. [42]Another centralizing factor is that these land uses are "communication-oriented" and generate heavy information flows among themselves. See E. M. Hoover and R. Vernon, Anatomy of a Metropolis (Doubleday Anchor, Garden City, N.Y., 1962)~~, pages 59ff.~~ But this factor lies outside the Thünen framework of this chapter.

---

969 ¶ 43. [43]For further discussion, see ~~my~~ Dalen's Essays in Spatial Economics, ~~pages~~ pp. 179–182.

---

970 ¶ 44. [44]See G. Sjoberg, The Preindustrial City (Free Press, Glencoe, Ill., 1960), ~~pages~~ pp. 97–100; L. F. Schnore, "On the Spatial Structure of Cities in the Two Americas", ~~Chapter 10 of~~ in The Study of Urbanization, P. M. Hauser and L. F. Schnore, eds., ~~editors~~ (Wiley, New York, 1965), ch. 10.

---

970 ¶ 45. [45]Examining urbanized areas in the 1960 U.S. Census, Schnore finds that the younger central cities tend to have higher status residents than their suburbs, while the older

central cities have the opposite, more usual, pattern. This may reflect a housing quality effect, the younger cities not having had time to become too dilapidated toward the center. L. F. Schnore, "The Socio-Economic Status of Cities and Suburbs," American Sociological Review, 28:76–85, (~~February,~~ 1963).

---

977 ¶46. [46] J. F. Kain, "Urban Travel Behavior", ~~pages 161-192 of~~ [in] Urban Research and Policy Planning, L. F. Schnore and H. Fagin, [eds.,] ~~editors~~ (Sage Publications, Beverly Hills, Cal., 1967), [pp. 161–92,] ~~especially pages~~ [pp.] 173–175.

---

979 ¶47. [47] Sjoberg mentions external danger as one factor making for centralized residences for the rich in the "preindustrial" city. ~~G.~~ Sjoberg, ~~The~~ Preindustrial City ~~(Free Press, Glencoe, Ill., 1960), pages~~ [pp.] 97–100.

---

980 ¶48. [48] Reasons for this decline are given in A. Lösch, The Economics of Location, [trans.] W. H. Woglom and W. F. Stolper, ~~translators~~ (Yale Univ~~ersity~~ Press, New Haven, 1954), ~~page~~ [p.] 43, n~~ote~~ 10.

---

983 ¶49. [49] W. Isard, Location and Space-Economy (M.I.T. Press, Cambridge, 1956), App~~endix~~ to Chap~~ter~~ 8. A notable pioneering effort in this direction, rich in descriptive detail, is that of R. M. Hurd, Principles of City Land Values [1st ed.] (The Record and Guide, New York, ~~1st edition,~~ 1903).

---

983 ¶ 50. [50]E. W. Burgess, "Urban Areas", ~~pages 113-138 of~~ in Chicago: An Experiment in Social Science Research, T. V. Smith and L. D. White, eds. ~~White, editors~~ (~~University of~~ Chicago Press, Chicago, 1929), pp. 113–138. The theory first appeared in 1925.

984 ¶ 51. [51]~~e.g.~~ For example, Illinois Department of Public Works and Buildings, Chicago Area Transportation Study, 3 vols. (Western Engraving and Embossing Company, Chicago, 1960–1962).

984 ¶ 52. [52]D. Rogers, 110 Livingston Street (Random House, New York, 1968), page 44.

984 ¶ 53. [53]~~e.g.~~ For example, N. Spulber, The State and Economic Development in Eastern Europe (Random House, New York, 1966), ~~pages~~ pp. 17–18 on Budapest; ~~A.~~ Lösch, ~~The~~ Economics of Location ~~(Yale University Press, New Haven, 1954)~~, page 129, note 4, on Vienna and Prague.

984 ¶ 54. [54]L. F. Schnore and D. Varley, "Some Concomitants of Metropolitan Size," American Sociological Review, 20 (Aug. 1955): 408–414, ~~August, 1955,~~ note the tendency for large cities to have water access.

985 ¶ 55. [55]E. F. Frazier, "Negro Harlem: an Ecological Study", American Journal of Sociology, 43:72–88, (July, 1937). He even pinpointed the nucleus: at the intersection of 7th Avenue and 135th Street.

985 [56] On farms and farm villages as Thünen systems see M. Chisholm, Rural Settlement and Land Use (Hutchinson University Library, London, 1962), chapter 4, and A. Lösch, The Economics of Location (Yale University Press, New Haven, 1954), page 62 note 45.

---

987 [57] See E. L. Ullman, "Regional Development and the Geography of Concentration," Regional Science Association Papers, 4:179-198, 1958, and C. D. Harris, "The Market as a Factor in the Localization of Industry in the United States", Annals of the Association of American Geographers, 44:315-348, December, (1954) 1954.

---

987 [58] Economic Commission for Europe, Economic Survey of Europe in 1954 (Geneva, 1955), chapter 6. See also A. Melamid, "Some Applications of Thünen's Model in Regional Analysis of Economic Growth," Regional Science Association Papers, 1:L1-L5, 1955.

---

gal 10 989 [59] For maps, and correlations with various social phenomena, see inter alia J. Q. Stewart, "Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population," Geographical Review, 37:461-485, July, 1947; J. Q. Stewart, "Demographic Gravitation: Evidence and Applications", Sociometry, 11:31-58, February-May, 1948; J. Q. Stewart and W. Warntz, "Macrogeography and Social Science," Geographical Review, 48: (Apr. 1958) 167-184, April, 1958; W. Warntz, Macrogeography and Income Fronts (Regional Science Research Institute, Philadelphia, 1965).
Localization of industry,
the market as a factor
C. D. Harris, cited above, also uses various potentials for market delineation.

Warntz,

(990) 60. ~~60~~ Macrogeography ~~and Income Fronts, cited above~~, cover and p~~age~~ 92. His map of world population-potential, p~~age~~ 111, has peaks in China and India, as one might expect, and conforms much more poorly to the world trade and travel pattern.

---

(991) 61. ~~61~~ Quoted in Barbara Ward, Spaceship Earth (Columbia Univ~~ersity~~ Press, New York, 1966), p~~age~~ 130.

---

(992) 62. ~~62~~ G. T. Trewartha, A Geography of Population: World Patterns (Wiley, New York, 1969), p~~age~~ 79.

---

(994) 63. ~~63~~ For more detailed discussion of these concepts, see ~~my~~ Toden Essays ~~in Spatial Economics~~, ~~pages~~ pp. 20–2~~4~~3.

---

(997) 64. ~~64~~ For descriptions see H. S. Perloff, E. S. Dunn, Jr., E. E. Lampard, and R. F. Muth, Regions, Resources, and Economic Growth (~~The~~ Johns Hopkins Press, Baltimore, 1960); R. A. Billington, Westward Expansion (3rd ed. (Macmillan, New York, ~~3rd edition~~, 1967); G. A. Lensen, ed~~itor~~, Russia's Eastward Expansion (Prentice-Hall, Englewood Cliffs, N.J., 1964).

In The Great Frontier (Houghton Mifflin, Boston, 1952), Walter Prescott Webb treats the expansion of Europe (the Metropolis) since 1500 into the newly discovered lands (the Frontier) as the central theme of modern history. Curio~~s~~usly, though Russia is part of the Metropolis, Siberia is not part of the Frontier (map, p~~age~~ 10).

[65] But see I. Bowman, The Pioneer Fringe (American Geographical Society Special Publication No. 13, New York, 1931), on the changing character of modern pioneering.

[66] W. A. Lewis, "Economic Development with Unlimited Supplies of Labour", Manchester School, 22:139-192, May, 1954. For more recent work see J. C. H. Fei and G. Ranis, Development of the Labor Surplus Economy (Irwin, Homewood, Ill., 1964).

F. J. Turner,

[67] The Frontier in American History (Holt, New York, 1921). For more recent work, much of it critical, see The Turner Thesis, G. R. Taylor, ~~editor~~ (Heath, Boston, 1956). Webb's Great Frontier, cited above, applies this thesis to the entire Western world.

---

[68] See ~~W. R.~~ Thompson, ~~A~~ Preface to Urban Economics ~~(The Johns Hopkins Press, Baltimore, 1965),~~ chapter 1.

---

[69] The pioneering work is W. Christaller, Central Places in Southern Germany, C. W. Baskin, translator (Prentice-Hall, Englewood Cliffs, N.J., 1966), published in 1933. Much of the literature is reviewed in ~~B. J. L.~~ Berry and ~~A.~~ Pred, Central Place Studies ~~(Regional Science Research Institute, Philadelphia, 1961).~~ (The second printing has a supplement through 1964). B. J. L. Berry, Geography of Market Centers and Retail Distribution (Prentice-Hall, Englewood Cliffs, N.J., 1967), is a good recent work in this tradition. An interesting verbal synthesis of the ideas of Thünen and Christaller, yielding a hierarchical arrangement similar to the one described here, may be found in E. von Böventer, "Towards a Unified Theory of Spatial Economic Structure," Regional Science Association Papers, 10:163-187, (1962), at pages 184-186.

---

[70] S. Kuznets, "Quantitative Aspects of the Economic Growth of Nations, X, Level and Structure of Foreign Trade: Long-Term Trends," Economic Development and Cultural Change, vol. 15, No. 2, Part II, January, 1967, pages 3-7.

1005 ¶71. [71]The central theme of N. S. B. Gras, <u>An Introduction to Economic History</u> (Harper and Brothers, New York, 1922), is the successive rise to dominance of larger and larger centers and associated hinterlands, i.e. ~~- that is~~, of Thünen systems, ~~-~~ from villages to world metropolises.

1005 ¶72. [72]The kind of dynamic Thünen analysis we have been discussing ~~in this section~~ has much in common with the literature on "growth poles" or "growth centers". See, e.g. ~~for example~~, J. Friedmann, <u>Regional Development Policy</u> (MIT Press, Cambridge, 1966), ~~and~~ J. R. Boudeville, <u>Problems of Regional Economic Planning</u> (Edinburgh Univ~~ersity~~. Press, Edinburgh, 1966). A very incisive analysis is to be found in A. O. Hirschman, <u>The Strategy of Economic Development</u> (Yale Univ~~ersity~~. Press, New Haven, 1958), ~~chapter~~ ch. 10.
