Towards a general theory of access

David Levinson  Hao Wu
University of Sydney  University of Sydney

Abstract: This paper integrates and extends many of the concepts of accessibility deriving from Hansen’s (1959) seminal paper, and develops a theory of access that generalizes from the particular measures of access that have become increasingly common. Access is now measured for a particular place by a particular mode for a particular purpose at a particular time in a particular year. General access is derived as a theoretical ideal that would be measured for all places, all modes, all purposes, at all times, over the lifecycle of a project. It is posited that more general access measures better explain spatial location phenomena.

Article History:
Received: August 25, 2019
Revised version submitted: March 4, 2020
Accepted: March 9, 2020
Available online: June 12, 2020

1 Introduction

The only reason to locate anywhere is to be near some people, places, and things, be far from others, and possess still others. Since being far from something is really just being near the absence of that thing, and possession is just the ability to have something (and legally prohibit someone else from having it), we can see that location is about proximity. People make location decisions all the time, from whether to move from North America to Australia, to traveling to the mall by car or bus, to standing near a person at a reception, or even sitting on the chair or the couch.

Firms do likewise, from where to build a factory, to where to locate a store, to which shelf to put the soda on the supermarket aisle to maximize profits (Underhill 2004, 2009).

The underlying logic of all these decisions is the same, despite the difference in scale, timeframe, motivation, and mode of travel. We want to understand what it means to measure general access to all the people, places, and things that people care about.

In practice, measuring everything is impossible, and so as empiricists we have limited ourselves to measuring things that are both important and feasible to measure. In this paper, however, we will see how far we can push the idea of access to everything. (The terms access and accessibility are used interchangeably.)
The evidence we have for access explaining location decisions is vast. We list some of the key observations which are largely explained by access here, with a few of the many references supporting the relationship. Access explains much of the variation in: Commuting Time (Levinson 1998; Vandersmissen et al. 2003; Wang 2000), Employment Rates (Ihlanfeldt and Sjoquist 1990; O’Regan and Quigley 1999; Shen 1998), Mode Shares (Kockelman 1997; Moniruzzaman and Páez 2012; Owen and Levinson 2015), Real Estate Prices and Density (Iacono and Levinson 2017; Shin et al. 2007; Srour et al. 2002), Incomes and Productivity (Cervero 2001; Deng 2013; Melo et al. 2017), and Investment Decisions (Levinson 2007; Levinson et al. 2015).

Cities exist because people want to be near each other, so residents and workers don’t have to travel as far to engage in routine activity, possessing what economists call economies of agglomeration. Physical infrastructure networks exist to connect within and between places faster than travel over the unimproved landscape. While travel on a network may be less direct (longer distance, more circuitous) than traveling ‘as the crow flies,’ it is also faster, or else why do it. People pay a premium to be in locations with greater access to the things (people, opportunities) they care about, to save time and spend less on travel, and to be more productive and earn more, all else equal.

The use of access to measure the performance of cities and networks has grown in recent years. While mobility measures the ease of moving on the network, accessibility measures the ease of reaching valued destinations. It thus considers both mobility on the network and the spatial distribution of the places people want to visit.

We refer to access in its geographical sense, though it has related meanings in the disability community: for instance, continuing with transport examples, is a bus accessible to someone in a wheelchair, or is the user interface on the walk signal actuator accessible for someone with poor vision?

With the explosion of big data, in particular continuously measured travel times from global positioning system (GPS) units in mobile phones and vehicles, automatic vehicle location (AVL) systems, standardized transit network representations in the standardized General Transit Feed Specification (GTFS) format, the OpenStreetMap project, satellite imagery, among others, all capable of being systematically mapped in geographic information systems, it is increasingly feasible to observe, measure, and monitor access in real-time. Coupled with increasingly precise computations from travel models which have long been available, and far more detailed digital representations of cities, access is becoming a standard output of transport applications, and should be a standard input into transport decision-making.

The work is titled Towards a General Theory of Access. There are three main words there: General, rather than specific, means we want to explain this broadly, not just under special circumstances; Theory, rather than practice, means we are interested as much in the idea as in operational questions. There is a forthcoming Transport Accessibility Manual which considers how to operationalize measures of access; and Access, rather than anything else.

Theory has different meanings to different people. We adopt the National Academy of Sciences (1999) definition:

“a well-substantiated explanation of some aspect of the natural world, based on a body of facts that have been repeatedly confirmed through observation and experiment.”

So there are two main aspects:

• Our general theory of access must explain a feature of the natural world (here we take the natural world to be location behavior).
• Our theory must be consistent with observations.

As noted above, there is a wealth of evidence that partial measures of access (access to particular destinations at particular times by particular modes measured in particular ways) explain much
Towards a general theory of access

Nevertheless, partial measures have come under increased scrutiny and criticism (Miller 2018). While we can never prove, in Popperian, falsificationist sense (Popper 1962), that a general and complete measure of access will explain everything about location, the general theory of access generates the hypothesis that a more general and complete measure of access will explain more about some location-based question than a less complete, more partial measure of access. Here we explore what a complete measure looks like.

To be clear, this is not a review paper, of which there are a few. (El-Geneidy and Levinson 2006; Geurs and Van Wee 2004; Niemeier 1997; Wu and Levinson 2020).

Although many ideas are covered and cited here, many are excluded as they follow different branches of understanding access. We are considering location-based access from a place, that is its classic geography sense. If the place is small enough, it uniquely represents a person at that place. A time-geography approach follows a person across space, and measures place-based access at different places across a time-space path, those different accesses at different places undoubtedly shape the choices individuals make across the day, and give insights by combining access in new ways.

A utility-based approach allows comparison of various places for various trip purposes and modes based on similar inputs, but weight those inputs based on a statistical estimation of what individual weights best explain a particular travel choice (e.g., mode, destination). A utility-based approach and location-based approach can be equivalent under specific assumptions (Wu and Levinson 2020), and offers an approach for combining the access provided by particular modes or across different activity types.

This paper provides systematic development of new ideas in what is referred to in the literature as location-based accessibility that follow from the explicit presentation of the underlying logic of general access in subsection 2.3. We lay these out in a building blocks manner both to systematize what we know, but also identify lacuna in our knowledge.

In the next sections we discuss what it means to measure general access. We discuss the different dimensions of access, starting with the question of ‘where’, (section 2), and define explicitly what we mean by access for one particularly useful and widely employed class of measurement primal access. We discuss impedances (section 3) and combining isochrones (section 4). We then proceed to systematically relax the embedded assumptions about time-of-day (section 5), activities (section 6), modes (section 7), and groups (section 8). We generalize the explicit representations of access (section 9). Finally, we extend that by considering access over the lifecycle of a project or place (section 10). We conclude (section 11) by identifying some gaps and issues for further development.

2 Where

We start with where (i). Where is the access being measured? We divide space into smaller units of geography (at the limit, every point; or every person, as per time-geography (Miller 1991); more practically for application, every parcel or building or street block or Census block group, or transport analysis zone), which we denote as (i). We then measure access (A_i) for that place.

2.1 Hansen’s Access

We solve for how many destination opportunities (O) (see subsection 6.1) can be reached from origin i constrained by some cost measure (f(C_{ij})) (see section 3).

This gives us what has been called the Hansen Accessibility measure after its developer (Hansen 1959). We refer to this as a Primal Access measure, to contrast with the Dual Access measure as described...
in Cui and Levinson (2019b), which considers the time or cost required to reach a fixed number of opportunities.

\[ A_i = \sum_{j=1}^{J} O_j f(C_{ij}) \]  

(1)

This primal measure of access is foremost positive, measuring how many activities can be reached. One could, however, impose a normative standard, and insist that it should be above some number (N). It implies the question of whether providing such levels of access is an appropriate role of government. For some activities, most people probably agree that it is (a fire station should be within X minutes of anyone who lives in a city, or X + Y minutes in a rural area), and for others it is not. As noted by Miller (2018), there is no standard for the number of jobs reachable within a threshold like 30 minutes, but all else equal, more is better than fewer. Discussions of the ‘30-minute city’ (Greater Sydney Commission 2018), for instance, suggest that 75% of the people should be able to reach jobs within 30 minutes using transit. Note, this definition bears on the distribution of jobs and housing as much as on public transport service.

2.2 Active and Passive Access

Equation 1 measures active accessibility. In contradistinction, how easy it is to be reached (the cost from j to i), rather than to reach (the cost from i to j), has been referred to as passive accessibility (Cascetta 2009; Papa and Coppola 2012), shown below:

\[ A^Q_i = \sum_{j=1}^{J} O_j f(C_{ji}) \]  

(2)

Passive accessibility has applications such as retailers or distributors who want to know how easily customers can reach them, or employers seeking to measure a prospective labor force, and may have relevance in understanding labor market competition (subsection 6.4) and access for delivered goods ordered online.

2.3 Explicit Access Formulation

Our explicit access \( (A_{i, h, z, m, e, t, p}) \) formulation expands Hansen’s first formulation given in Equation 1, solving for how many destination opportunities \( (O) \) can be reached from an origin \( (i) \), by an activity type or purpose \( (z) \), for a population subgroup \( (p) \) (income category, racial group, modal availability, etc.), with a given mode \( (m) \), at a time of day \( (h) \), for a given set of activities \( (z) \), considering cost elements \( (e) \), and an impedance threshold \( (t) \). While this may seem pedantic, it also reveals the richness of the problem, which is multi-dimensional, and opens avenues for generalization.

\[ A_{i, h, z, m, e, t, p} = \sum_{j=1}^{J} O_{j, h, z} f(C_{ij, h, m, e}) \]  

(3)

2.4 Person-weighting

Person-weighted access produces more representative measures than arithmetic averages, and reflects access as experienced by the population. It is a common practice when aggregating accessibility (Geurs et al. 2003; Wachs and Kumagai 1973).

To obtain an aggregate measure, we can average all of the accessibilities measured at all of the origin \( i \)'s, and develop a system-wide average. But how you average matters, so we often person-weight the average.
To obtain a system average, we might sum the access in each origin, weighted by the number of people experiencing that access, for instance those who live in that zone \( S_{i_c} \). We might go further, and consider daytime and nighttime populations, or tracking individuals through their personal time-space prism (Tang et al. 2016), but that is beyond the scope of this paper.

First we note the population \( S_{i_c} \) in a zone is the sum of all the subgroups \( p \) in the zone.

\[
S_{i_c} = \sum_{p=1}^{P} S_{i_c,p}
\]

We then give the person-weighted average of the regional access, weighted by the zonal population.

\[
A_{i,b,z,m,e,t,c} = \frac{\sum_{i=1}^{I} A_{i,b,z,m,e,t,p} S_{i_c}}{\sum_{i=1}^{I} S_{i_c}}
\]

3 Impedance

Time, distance, money cost, and other travel related expenses impede travel, which reduces access. Initially the question of how much \( C \) the trip between \( i \) and \( j \) cost was taken to be a function of distance, modeled on Newton’s Laws of Gravity. Later travel time was used, and monetary costs have been directly considered in a few analyses (Cui and Levinson 2019a; El-Geneidy et al. 2016). The travel impedance for the cumulative opportunity access measure is reflected in two forms:

- An impedance function \( f \) discounting the weighted number of opportunities based on the travel cost to such opportunities, discussed in subsection 4.1.
- A cost \( C_{ij} \) expressed by the travel cost to all, or for a given level of, opportunities reached.

Since there is a one-on-one correspondence between the travel cost from a point, and the number of opportunities reachable at that cost threshold, access can be expressed either by the cumulative number of opportunities reachable (primal access measure \( \text{section 4} \)) or alternatively by the travel cost (dual measure (Cui and Levinson 2019b)) for reaching a certain level of opportunities.

Travel impedance is incorporated into access by discounting cumulative opportunities by their travel cost. This section discusses constituents of the travel impedance.

3.1 Time Cost

In early applications, distance was used as a measure of travel cost, but now travel time is more widely used. The same time cost is experienced relatively equally by different demographic groups. The travel time of a route depends on the speed of travel along the network and distance of the route \( s_{ij} \), and so the efficiency (circuity) of the network design is often as important as the velocity \( v_{ij} \) on the network in determining travel time.

\[
t_{ij} = \frac{s_{ij,\text{network}}}{v_{ij}}
\]

Fluctuations in traffic conditions result in different shortest time paths by time-of-day (Moya-Gómez and García-Palomares 2015); transit travel time varies with schedules (Farber et al. 2014; Lei and Church 2010).

Time cost can flexibly combine multimodal, and different stages, of a trip. The travel time cost \( t_{ij,e,m} \) of a one-way trip by mode \( m \) can be measured as a linear summation of the various time elements \( e_i \) in the set of relevant time elements \( E_i \). For transit the relevant cost elements might be waiting, on-board, and transfer time penalties (Lee et al. 2013; Shen et al. 2014). Similarly, adding
walking time to and from parking lots to the automobile travel time (without weighting), helps travel
time between different modes becomes more comparable (Benenson et al. 2011).

\[ t_{ij,m} = \sum_{e_t=1}^{E_t} t_{ij,e_t,m} \]  \(7\)

For competing modes of transport, the optimal travel time \(t^*_{ij}\) can be represented by the fastest
mode between the respective locations, shown in Equation 8 (Shen et al. 2014).

\[ t^*_{ij,m} = \min_{m} (t_{ij,m}) \]  \(8\)

We discuss weighting time with a function to account for perception or user experience in subsection 3.2.

### 3.2 Perceived and Experienced Time

The cost that is used in access is generally estimated by the analyst, and is often assumed to be objective,
but we may want to think about how users perceive cost. Time perception varies with conditions under
which time is experienced, and while it varies from person to person, there are conditions which most
people will perceive to be longer than actual (waiting, stop and go traffic) and conditions which will
be perceived to be shorter (moving unimpeded).

While getting the perceived time for every origin-destination pair for every individual is likely to
be impossible, models of perceived time can be estimated and used as part of the cost matrix \((C_{ij})\).

Travel is generally assumed to be a ‘derived demand’, an undesirable process people endure to reach
something they desire. The perception of travel cost systematically deviates from objective measures
(Carrion and Levinson 2019; Fan et al. 2016; Lagune-Reutler et al. 2016; Parthasarathi et al. 2013; Wu et al. 2009). We talk about perceived and experienced cost, because even if a traveler has correct
time perceptions, they may still favor one travel experience over another. Different mode and trip pur-
poses produce different functions reflecting decay of trip rate with distance; active modes of transport
decay most rapidly with distance, as the perceived cost and physical exertion are much higher than
motorized modes (Iacono et al. 2008a). Notably, time is perceived and experienced differently for
different modes, and for different stages of a trip. Transit users perceive bus waiting time as longer
than the objective measure, and surveys find it to be between 15% to 40% longer than actual (Jones
and Peppiatt 1996; Larson et al. 1991; Watkins et al. 2011), and the extent of overestimation varies by
individual characteristics, such as age and gender (Jones and Peppiatt 1996). In light of the difference
between actual and perceived time, travel time at different stages of a trip can be assigned a weight \((\eta_{e_t})\) for each time element \((e_t)\) to produce a single perceived travel time \(t_{ij,m}^P\) that reflects the overall
travel experience, with on-board time generally less onerous than walking or transit waiting time
(Chang and Lee 2008; Tahmasbi and Haghsenas 2019).

\[ t_{ij,m}^P = \sum_{e_t=1}^{E_t} \eta_{e_t} t_{ij,e_t,m} \]  \(9\)

Related to perceived time is the idea of the quality of time. People may accurately perceive time,
but wish it were different. Handling this is more difficult. Modes vary in the quality of time. Different
modes require different amounts of effort or energy expenditure. People are happier walking or riding a
bike than riding a bus or crowded train or driving (Abou-Zeid et al. 2012; Chen et al. 2019; Stutzer and
Frey 2008; van Lierop and El-Geneidy 2018). Within mode choices models, for transit, for instance,
the weighting on out-of-vehicle waiting time is much greater than on in-vehicle time, and while some
of this accounts for time perception, and some accounts for uncertainty (is the bus actually coming?),
some also accounts for time quality.
Towards a general theory of access

By using satisfaction as a modifier to the travel cost, jobs could be discounted according to the degree of dissatisfaction associated to the trip by a given mode (Chaloux et al. 2019). For example, two jobs reachable with a trip satisfaction rating of 50% may be worth one job reachable with a satisfaction rating of 100%. This would enable the use of satisfaction-based access in an easily-understood cumulative opportunities framework. Combining perceived (or reported) travel time (as distinct from objectively measured travel time using GPS) with satisfaction in access measures is another direction for future research. The issue, which cannot be addressed with current data sets, is the extent that dissatisfaction already embeds higher perceived travel times, or the degree to which they are two distinct phenomena.

3.3 Generalized (Internal) Cost

The generalized cost (not to be confused with generalized access (section 9)) combines two or more elements of the internal travel cost, the cost borne by the traveler, so the access measurement gets closer to reality. The cost elements included by the generalized cost are not as comprehensive as the full cost discussed in subsection 3.4, which includes externalities.

Thus far we have abstracted cost \( C_{ij} \). In most applications cost has been taken as individual travel time, so a primal accessibility measure might ask how many jobs a traveler can reach in 30 minutes of travel. While this is useful for many applications, it neglects many other costs of transport. From the user perspective, costs include monetary expenditures on travel, often proportional to travel distance (Ingram 1971; Richardson and Young 1982), travel time (Farber et al. 2014; Moya-Gómez and García-Palomares 2015; Osland and Thorsen 2008), tolls, transit fares, parking, fuel, costs of vehicle ownership, and so on. The cost of travel can be monetized (by converting time to money) or temporalized (by translating money costs to time), for instance by considering the amount of time required to work to earn enough to pay transit fares. (El-Geneidy et al. 2016)

Time and monetary costs are the most often used combination with the generalized cost. The travel time can be monetized \( C_{ij,b,m,t} \) using a conversion factor \( w \) that is the value of time or the wage rate. (El-Geneidy et al. 2016; Goodwin 1974)

\[
C_{ij,b,m,t} = w \cdot t_{ij,m}
\]  

Similarly monetary cost can be denominated with time units \( C_{ij,b,m,t} \):

\[
C_{ij,b,m,t} = \frac{S_{ij}}{w}
\]  

Hence the generalized internal cost function can account for both time and monetary costs experienced by users while expressing spatial friction in a single interpretable unit (either time or money).

\[
C_{ij,b,m,int} = \sum_{e_{int}=1}^{E_{int}} C_{ij,b,m,e_{int}}
\]

The actual monetary value of time has been found to be a percentage (usually less than 100%) of the corresponding wage rate for the same length of time, and the value differs between driving alone and carpooling (Lam and Small 2001). Higher income groups tend to accept higher monetary cost in exchange for shorter travel time, so having higher monetary value of time. (Beesley 1965; Hensher 2019) There is a large literature on appropriate values of time, which we cannot fully review here. The idea of generalized cost, and converting time into money or vice versa, is important not only in estimating the travel cost, but in considering cost thresholds as discussed in 4.1.
3.4 Full Cost

In social welfare analyses, the travel time and the monetary costs paid by travelers are only part of the cost for making a trip. Significantly, someone incurs the cost of providing subsidy for those costs the travelers do not bear, for instance when the financial cost of providing the service exceeds the revenue it generates from users. But even more, there are externalities, like pollution, noise, crashes, and congestion imposed on others which should be considered, as they are borne by society at large.

From society’s perspective, welfare is maximized not when only user costs are considered but when society’s full cost are. If access is to be used in evaluation, it must consider these factors. In this case, we need to consider other cost elements ($e_{full}$), like congestion imposed on others, pollution emitted from the vehicle, danger from crash risk, noise, and infrastructure and other subsidies provided to travelers. For the same cost threshold, inclusion of external costs reduces the number of opportunities reachable compared to using only the internal travel cost (Cui and Levinson 2018). Full cost access has significant policy implications, as the optimization process (if optimizing for access) shifts from individual to system optimum; the inclusion of externalities remedies the inherent market failure, and should lead to a more efficient market; the distribution of the newly introduced fare burden has equity implications.

$$C_{i:j;h:m:e_{full}} = \sum_{e_{full}=1}^{E_{full}} C_{i:j;h:m:e_{full}}$$  \hspace{1cm} (13)

Thus we can compare the generalized (or internal) and full (or internal plus external) costs of travel.

Our hypothesis is that while automobile is often faster than other modes (and so has the highest time-based access measure), it is unlikely to have the lowest full cost of travel in urban areas.

Inclusion of trip externalities (which impose safety, health and environmental cost on other users, non-users) into the travel cost function produces different route costs. Accessibility outcomes depend on the choice of paths (safest, cheapest, greenest, and full cost paths found by using crash risk, monetary, environmental and full costs). This logically affects the shape and size of the isochrone catchment area, and the number of opportunities reachable (Cui and Levinson 2018).

Conceiving of the problem with the ‘full cost’ of access changes the perspective from individual to the system optimum, which has policy implications, for example in the possibility of charging user fees to internalize the externalities of travel, which would result in a more efficient use of the system. Different policy objectives can be pursued by selectively incorporating and weighting components of cost externalities.

We note that when estimating costs, the path that travelers traverse is usually assumed rather than observed. Often it is assumed travelers choose the shortest path. The evidence is however travelers do not, for reasons relating to time perception, among others (Tang and Levinson 2018; Zhu and Levinson 2015). The total cost depends very much on the behaviorally-based path choice of the traveler. If trying to understand full cost, or future scenarios where there might be significant road pricing, and travelers have different values of time, these technical details may significantly affect outcomes.

4 How We Assess

In Hansen’s access (subsection 2.1), we might use an impedance function ($f$) with a travel time threshold ($t$) or with specific parameters ($\theta$), which give different, though systematically related, numeric answers.

In this section we connect the various assessment methods for primal access.
4.1 The Isochrone

The isochrone, from the Greek word for \textit{same time} is the easiest window into access. An isochrone map shows for a point, what places can be reached in a given window of cost, usually measured as time, for instance 10 minutes, 20 minutes, so on. The cost can be measured in money or other terms as well.

Isochrone access measures are operationalized by the summation of the number of opportunities encapsulated within a predetermined travel cost threshold (usually time or distance) (El-Geneidy and Levinson 2007). The measure has a geometrical explanation: an isochrone represents all possible destinations within the travel cost threshold, opportunities within the catchment area of the isochrone are considered reachable and those outside are discounted (Ingram 1971; Sarlas and Axhausen 2018). The isochrone method is a special case in which the decay function is rectangular. For its ‘all-or-nothing’ approach in counting potential opportunities, the isochrone measure is sometimes called a cut-off, contour, cumulative opportunities, proximity or cumulative count (Neutens 2015).

The actual route choice tends to be more circuitous than necessary, given that people do not have perfect information about the road network, and may not wish to simply minimize travel time (Tang and Levinson 2018; Zhu and Levinson 2015). Hence an isochrone estimated via an analyst’s shortest time path likely represents a potential coverage area under ‘ideal’ circumstances. The actual area, and number of opportunities, reachable by most people tends to be smaller than what is covered by the isochrone.

The operationalized cumulative opportunity access measure uses the isochrone as its (rectangular) decay function.

To apply access measures in practice, the function of costs \( f(C_{ij,h,m,e}) \) needs to be specified. For completeness, here we note the subscript for origin \((i)\), destination \((j)\), time-of-day \((h)\), and for mode \((m)\), considering a particular set of costs \((e; \text{time/monetary})\), and recognize that we could add any other sub-categorization we may wish to impose.

We call this a rectangular distance decay function as it provides a step function for inclusion or exclusion of an opportunity from the access set. The formulation of the travel impedance function demarcating the isochrone is given by Equation 14. Tolerance for travel cost differs by mode, (Iacono et al. 2010; Levinson 1998), trip purpose, (Grengs 2015; Hansen 1959), and cost elements considered, so the upper bound travel cost threshold \((t_{ub})\) would vary accordingly, which we could denote as \((t_{m,z,e})\), but don’t develop that further here.

\[
f(C_{ij,h,m,e}) = \begin{cases} 
1 & \text{if } C_{ij,h,m,e} < t_{ub} \\
0 & \text{otherwise} 
\end{cases}
\]  

(14)

4.2 Travel Time Budgets

A behavioral foundation justifying the use of cumulative opportunities hinges on the constancy of the travel cost budget, which determines the spatial extent of the isochrone coverage area. The individual commute duration has been relatively stable over long periods of economic and demographic change (both with and without major technological change), suggesting a latent travel time budget from daily activities (Levinson and Kumar 1994; Marchetti 1994; Zahavi 1974). There has been mixed evidence as whether travel time or monetary cost alone is constant; (Mokhtarian and Chen 2004), or whether the budget applies to commute trips or all travel, however the generalized travel cost as a summation of time and monetary expenditure, and expressed in time unit (generalized time) appear to be stable for short periods of time; when expressed in monetary units, the generalized cost is a stable proportion of individual income. (Goodwin 1974, 1981; Tanner 1981) This implies a trade-off between travel time and money expenditure that keeps the total generalized cost stable. Research suggests time budgets for public transport commuters may be higher (about 45 minutes each way) (Huang et al. 2018; Wu et al. 2020, 2019) than for autos (about 30 minutes each way).
Redmond and Mokhtarian (2001) suggest a positive utility of commute, so that travelers who want to travel a minimum amount in addition to being less than a travel time budget (Mokhtarian and Chen 2004). We can express this with time as in Equation 15, where \( t_{lb} \) is the lower bound and \( t_{ub} \) is the upper bound of the time threshold as before. The lower bound makes less sense for a monetary cost threshold (\( C_{lb} \)), but one could impose a non-linear value of time, while the upper bound (\( C_{ub} \)) is straight-forward.

\[
f(C_{ij,b,m,e}) = \begin{cases} 
1 & \text{if } t_{lb} < C_{ij,b,m,e} < t_{ub} \\
0 & \text{otherwise}
\end{cases}
\] (15)

The isochrone measure counts the actual number of opportunities reachable, and has the advantage of being in absolute and comparable units (Batty 2009). The measure provides better interpretability than using a decay function and the resulting relative measure. The clarity of meaning and easily understood and interpretable concept makes isochrone access ideal for practice (O’Sullivan et al. 2000).

The use of various travel cost thresholds inherently produce different cumulative opportunity measures, and more importantly may change the rankings of alternatives of places in a relative sense (Miller 2018; Owen and Levinson 2015; Xi et al. 2018). This presents a major challenge with the isochrone measure, since the selection of travel cost threshold significantly impacts its interpretation. While a large threshold (about 60 minutes by car) will include all destinations in most metropolitan areas, and thus be indicative of city size, a very small threshold (about 15 minutes) may capture relatively few jobs, and large cities may have lower person-weighted accessibilities than smaller places.

In practice, isochrone access is often measured in minute-by-minute increments, each a time annulus, e.g., the number of opportunities available between 13 and 14 minutes, and models using access as input to predict mode share, population density, etc. often take the travel time cut-off value that produces the best model fit (Owen and Levinson 2015; Xi et al. 2018). Isochrone access permits the calculation of a weighted sum of opportunities within incremental travel cost thresholds (Kwan 1998), and the best fit weighing function can be applied, which is similar to the idea of the distance-decay function.

4.3 Time-weighted Cumulative Opportunities

In a time-weighted cumulative opportunities (alternately called: gravity, potential, distance decay, or cost-weighted) formulation opportunities are weighted based on their travel cost from the place of interest. Distant opportunities, and opportunities with greater travel cost, tend to be weighted lower than nearer or less expensive opportunities reachable from the same location. When calibrated empirically, the impedance function reflects the diminishing trip rate with increasing travel cost.

Common forms of decay functions, including exponential, exponential-normal (Gaussian), exponential-square root, log-normal, log-logistic and negative power are listed in Table 1. The impedance functions are sometimes referred to as time or distance decay, and are often calibrated to match the decreasing trip rates with increasing travel cost from travel surveys. Different trip purposes (Grengs 2015; Iacono et al. 2008b), social demographics, level of education received (Geurs et al. 2003), and modes of travel affect the parameters of the impedance (Geurs et al. 2003; Tahmasbi et al. 2019).

One commonly used function is negative exponential. In this case we might write access as: \( A_{i,j,b,z,m,e,b,p} \) to account for the use of the impedance factor (\( \theta \)) rather than a time threshold (\( t \)).

The theoretical basis for using the exponential form is provided by the entropy maximization principle applied to the distribution of spatial interaction. The derived functional form of the impedance function is exponential (i.e. using Euler’s number (e), denoted as \( e^x \) to avoid confusion). (Hayes and Wilson 1971) This is functionally equivalent to utility in logit choice models.

\[
f(C_{ij,b,m,e}) = e^{x \theta_{z,e,m} C_{ij,b,m,e}}
\] (16)
Towards a general theory of access

Table 1: Illustrative Distance Decay Functions, Applicable to any mode \((m)\)

<table>
<thead>
<tr>
<th>Decay Function</th>
<th>(f(C_{ij}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>(\exp(\theta_e C_{ij}))</td>
</tr>
<tr>
<td>Gaussian</td>
<td>(\exp(\theta_e C_{ij}^2))</td>
</tr>
<tr>
<td>Log-normal</td>
<td>(\exp(\theta_e + \ln C_{ij}))</td>
</tr>
<tr>
<td>Exponential-square-root</td>
<td>(\exp(\theta_e C_{ij}^{0.5}))</td>
</tr>
<tr>
<td>Gravity</td>
<td>(C_{ij}^{-2})</td>
</tr>
<tr>
<td>Potential</td>
<td>(C_{ij}^{-1})</td>
</tr>
<tr>
<td>Scaling</td>
<td>(C_{ij}^{0.1})</td>
</tr>
<tr>
<td>No Distance Decay</td>
<td>1</td>
</tr>
<tr>
<td>Linear</td>
<td>((2 \cdot C_{ak} \cdot (C_{ak} + C_{ij})))(^{11})</td>
</tr>
<tr>
<td>Spline</td>
<td>(\theta C_{ij})</td>
</tr>
</tbody>
</table>

The impedance factor \((\theta_{m,z,e} < 0)\) is measured empirically and varies with mode \((m)\), activity or purpose \((z)\) and cost categories \((e)\). Previous studies have found values on the order of \(\theta = -0.08\) for work trips by car considering travel time. (Levinson and Kumar 1995) There are numerous other commonly used impedance functions, with empirically estimated parameters, as shown in Table 1. While \(C_{ij}\) is often taken to be the time-cost, it could be the distance cost, or any other cost.

How travel cost is expressed within the impedance functions can also be flexible, for instance, a polynomial of the same cost variable as in Equation 17.

\[
f(C_{ij},h,m,e) = \exp(\theta_0 + \theta_1 \sqrt{C_{ij},h,m,e} + \theta_2 C_{ij},h,m,e)\quad(17)
\]

A combination of different categories of travel cost including travel time, distance and monetary cost can be included as the generalized cost of the impedance function. (Ford et al. 2015) The functional form of the exponential decay function makes it easier to combine different categories of travel costs into a single generalized cost function; Equation 18 gives an example of combining travel time and monetary cost using an exponential decay function. (Bocarejo and Oviedo 2012)

\[
f(C_{ij},h,m,int) = \exp(\theta w \cdot C_{ij},h,m,int + C_{ij},h,m,s)\quad(18)
\]

\(C_{ij},h,m,int\): generalized (internal) travel cost between \(i\) and \(j\)

\(C_{ij},h,m,int\): time cost

\(C_{ij},h,m,s\): monetary cost

\(w\): wage rate or value of time.

5 When

Next we can consider when \((h)\). Accessibility problems are typically laid out as if all opportunities were available 24 hours a day. But access to opportunities varies over time. For instance access in the peak hour differs from that at 4:00 am. There are two reasons.

First, transport services vary by time-of-day (from minute-to-minute and hour-to-hour). The network differs at 4:00 am from that at 8:00 am. For roads there is less traffic at 4:00, and so roads might be faster. In contrast, for transit, there is less service at 4:00 am, and so more waiting or access/egress time. We also note that access at 8:00 am differs from 8:01 as well. Traffic differs somewhat due to the ebb and flow of congestion and shockwaves. Transit varies more systematically: at 8:01, the scheduled 8:00 am bus or train may have just left, increasing the waiting time at the stop or station, which implies the transit travel time differs greatly. Averaging transit travel times across a peak hour can provide a more realistic measure than just a sample at a single point in time (Owen and Levinson 2015; Owen and Murphy 2019). We may do a simple average, or something more complicated, as in subsection 2.4.
Second, the opportunities \((O_{i,h})\) differ by time-of-day. Stores and restaurants open and close. Jobs have hours when they are available. Further, the value of an opportunity, and thus overall access, depends on when it is experienced. People are more interested in certain places at certain times of day.

The idea of the space-time prism, or network-time prism is relevant to the questions of where and when. Every individual has a different network-time prism based on the opportunities available to them at different times of the day from the locations they are at then, and their schedule constraints (time-of-day varying time budget) (Song et al. 2017).

There is another scale of \(When\) which refers to longer periods of time. A traveler’s desire of what to access depends on when you are asking. The desires for a weekday may include their workplace, that may be irrelevant on the weekend.

The points of interest that are of interest in the summer may differ from those in winter; for instance in western countries, access to shopping is more essential to people in late November and December before the holidays than March or April.

The places a school-goer values differs from those a retiree considers. We discuss how to combine different preferences over the longer run in section 10.

We need to consider the cost of travel (including the travel time) at a given time-of-day \((h)\) by mode \((m)\) considering cost elements \((e)\) \((C_{ij,h,m,e})\). We may sum over opportunities available at a given time-of-day \((O_{j,h,z})\) if we appropriately weight them. Solving separately for a given trip purpose \((z)\) and mode \((m)\) at a given time-of-day \((h)\) and summarizing across the whole day gives:

\[
A_{i,,z,m,e,t,p} = \sum_{j=1}^{I} \sum_{h=1}^{H} U_h O_{j,h,z} f(C_{ij,h,m,e}) \tag{19}
\]

where \(h\) indexes time-of-day, and \(U_h\) weights the value of each time slice.

We normalize so that:

\[
\sum_{h=1}^{H} U_h = 1 \tag{20}
\]

We might choose \(U_h\) to be the fraction of the time-of-day represented by a time slice \((h)\). We can derive the minute-by-minute-averaged access over the course of a whole day (Farber and Fu 2017). With this scheme, access measured at each minute contributes equally towards the aggregated daily access; the weight \((U_h)\) of each time slice equals \(\frac{1}{1440}\). Other partitions of the time slices, and varying time spans for averaging access may be used, although the weight of all time slices must sum to 1, as required by Equation 20.

Time slices could also be estimated statistically using results collected from a survey.

A major motivation for time-weighting access is to smooth out its temporal fluctuations. Trip departure time affects travel cost especially when the network is congested (Bruinsma and Rietveld 1998); the level of access can be especially susceptible to trip departure time due to the temporal variation in transit service provision (Boisjoly and El-Geneidy 2016; Farber and Fu 2017). Time-averaged transit access better explains mode choice behavior than peak access (Owen and Levinson 2015).

6 Why

Examining \(why\) we travel (or \(what\) is being accessed) \((z)\) brings to the fore the question of what kind of opportunities are of interest. They may be jobs, or houses, or the number of jobs available after controlling for workers (see subsection 6.4), or stores, and so on.

Historically, when measuring access to jobs, analysts have considered the existence of jobs, and measured them as if jobs were interchangeable (though noting, of course, that jobs are not). Some have stratified jobs by income or by type to examine the number of jobs available to individuals with specific skills. This analysis depends upon data availability. Also not all jobs are available at the same
time. For instance, though both work in office buildings, cleaning crews tend to operate outside of the
regular business hours for desk workers, so this relates back to the question of *when*.

We should also note that perceived accessible opportunities (both perceived as an alternative oppor-
tunity, and reachable given schedule constraints) differ from the analyst’s measure of opportunities. (*Cascetta et al. 2016*).

### 6.1 Multiple Activities

The number of opportunities at each potential destination \((O_j)\) reflects its value for spatial interaction
to the traveler. This opportunities count is synonymous to the measure of destination attraction from
spatial interaction models. Destination attraction may be measured, for instance, with its population
(*O’Kelly and Horner 2003*), the number of job opportunities, GDP, intensity of specific activities
(*Bruinsma and Rietveld 1998*), size of the facility (*Knox 1978*), market potential or sales (*Holl 2007;
Spence and Linneker 1994*), or pixels of built-up area (*Vermeiren et al. 2012*).

The problems above are laid out as if there were only one opportunity or activity type or purpose of
interest, e.g., jobs. But the value of a location, and thus overall access, depends on many different types
of opportunities which one may reach from a location. An access measure involving multiple destina-
tion types can be weighted to produce a single aggregated measure, (*Hou et al. 2019*) or alternatively,
can be stratified and discussed separately (*Van Wee et al. 2001*).

We may sum over different types of opportunities (*Bentlage et al. 2013*) if we appropriately weight
them:

\[
A_{i,h,,m,e,t,p} = \sum_{z=1}^{Z} \sum_{j=1}^{J} W_{z,p} O_{j,h,z} f(C_{i,j,m,e})
\]  

(21)

where \(z\) indexes activity types, and \(W_{z,p}\) weights the value of each activity type for each subgroup.

We normalize the weight so that:

\[
\sum_{z=1}^{Z} W_{z} = 1
\]  

(22)

### 6.2 Weighting Activities

#### 6.2.1 A Weighted Average

A first thought would be to choose \(W_{z}\) as the normalized duration \((D)\), share of daily out-of-home
time, at each activity, or perhaps frequency of the activity. (*Hou et al. 2019*)

\[
W_{z} = \frac{D_{z}}{\sum_{z=1}^{Z} D_{z}}
\]  

(23)

We might prefer some other indicator of an activity’s importance. For instance, toilet use does
not typically comprise many minutes per day, but is often urgent, and the substitutes are socially un-
derirable. Similarly medical treatment may be important, but people would pay a premium to be in
a hospital for less time. This could be defined uniquely for each individual or subgroup \((p)\) if the
data were available. In network design problems using access as the objective function (maximizing
access in linear programming problems), the weighting scheme can be based on the subjective priority
assigned to each zone. (*Antunes et al. 2003*)

Another issue is the compatibility of counts, one job is not commensurate with one store, so the
matrix \(W_{z}\) should address that as well, and make the scales of opportunities consistent.
Example  Consider a family living in a zone with access to 3 jobs, 2 stores, 4 homes belonging to other people (excluding their own from access calculations) within 30 minutes. Each day, individual \( p \) spends 8 hours at work, 1 hour shopping, 1 hour visiting friends (in other people's houses), and 14 hours at their own home (of which there is 1). We neglect access to one's own home, though obviously including it would change the numeric answers. We take the sum over the amount of out-of-home time spent at activities.

We can denote the durations \( D_z \) and the opportunities \( O_z \) as:

\[
D_z = \begin{pmatrix} 8 \\ 1 \\ 1 \end{pmatrix}, \quad O_z = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}
\]

The access by the above definition (Equations 21, 22, 23) yields:

\[
\left( \frac{8}{10} \right) \cdot (3) + \left( \frac{1}{10} \right) \cdot (2) + \left( \frac{1}{10} \right) \cdot (4) = 3
\]

Doubling jobs (from 3 to 6) increases overall access sublinearly to:

\[
\left( \frac{8}{10} \right) \cdot (6) + \left( \frac{1}{10} \right) \cdot (2) + \left( \frac{1}{10} \right) \cdot (4) = 5.4
\]

6.2.2 A Cobb-Douglas approach

What we really have is a multi-factor production function, each activity is necessary for the production of full access. The Cobb-Douglas function was an early version of that, measuring output as a function of input (Douglas and Cobb 1928).

We can write that for our case as composite opportunities \( O_z \):  

\[
O_z = e^{W_0} \prod_{z=1}^{Z} (O_{j,h,z})^{W_z}
\]

or equivalently

\[
\ln(O_z) = W_0 + \sum_{z=1}^{Z} W_z \ln(O_{j,h,z})
\]

Where the coefficients in Equation 29 indicate increasing, constant, and decreasing returns to scale respectively:

\[
\sum_{z=1}^{Z} W_z > 1, \quad \sum_{z=1}^{Z} W_z = 1, \quad \sum_{z=1}^{Z} W_z < 1
\]

where the interpretation of a Cobb-Douglas model indicates that: \( W_0 \) is an efficiency parameter, and the weights \( (W_z) \) are elasticity parameters.

The weights could be assumed to be the fraction of out-of-home time spent at a given activity, or some other method could be employed to estimate their value, for instance based on their influence on real estate prices.

Example  Consider the same zone as above. Use share of daily out-of-time time as exponents, which has the notable property of ensuring constant returns to scale, as the shares sum to 1.

The access by the above definition (Equation 27) would be:

\[
O_z = 3^{8/10} \cdot 2^{1/10} \cdot 4^{1/10} = 2.96
\]
Towards a general theory of access

and doubling jobs without changing the other out-of-home activities increases access sublinearly, as expected.

\[ O_{,,..} = 6^{8/10} \cdot 2^{1/10} \cdot 4^{1/10} = 5.16 \]  

(31)

Qualitatively, this procedure feels more natural than a simple weighted average, though whether it better explains behavior is an open question. Obviously this example is dominated by jobs for two reasons, there are more jobs than other activities and more time is spent at jobs than other activities. So while the primal access to jobs would be 3 (or 6) in this example, the multi-activity access varies from 2.96 to 5.16.

6.2.3 Schedules and time-dependent activity weights

Our explication has largely been from the perspective of longer-term location decisions, and the opportunities available at places. We have not talked about daily schedules, and the choice of a location within the day (e.g., where to eat lunch) surely depends on what one plans to do next. So the weights for different purposes may themselves be time-of-day dependent. Exploring that awaits future research.

6.2.4 Discussion

The correct way of combining opportunities is far from obvious, even if we agreed jobs, stores, and houses were commensurate units. We also have the problem of which activities to consider, though presumably time spent at places helps inform this, and infrequent activities (like airport travel, the zoo, museums, or stadiums) don’t have the same importance as daily trips. Nevertheless, developing weights in the absence of human behavioral data (like travel choices or housing prices) is fraught. Using some form of statistical modeling (e.g., of house prices) to develop weights, or entering activities separately into (e.g., hedonic) models is another approach.

Using percentage of a region’s opportunities of type \( z \) available to normalize each type would help make the different activities commensurate in a given analysis, but makes it difficult to compare between places or over time. Someone in a small city (e.g., Fresno) may be able to reach 100% of opportunities in 45 minutes, while in a large city (e.g., Los Angeles), a smaller percentage of a much larger number can be reached in the same 45 minutes. Which has better access? We believe land markets primarily consider absolute rather than relative access when embedding value, recognizing that diminishing returns to access may make these more similar than constant returns to access would.

An assumption implicit in the approaches above is that opportunity types are substitutes. But clearly some are not, access to all may be needed for a successful place or person. A city with all stores, but no jobs cannot survive (nor can the opposite). So simple summation, with weighting, may be misleading.

Some activities are substitutes, or near substitutes (restaurants can substitute for grocery stores), while others are complements (people cannot buy food if they don’t have an income provided by jobs). In network terms, substitutes are parallel or competitive links, complements are links in series. The classic example in economics is shoes: left and right shoes are perfect complements, shoes and slippers are perfect substitutes. This is however complicated, as there are only 24 hours in a day, in a sense a minute at any activity competes with a minute at another activity. (Levinson and Kanchi 2002)

6.3 Diminishing Returns to Opportunities

In gravity models of access, discussed in subsection 4.3, the value of opportunity diminishes with the cost to reach it. But the value of opportunities may diminish simply because of their number. There is a limit to the available time per day, and the amount of service a person can reasonably consume. The addition of a second opportunity within 10 minutes adds less value than the first one, and so on. To account for this, we might wish to adjust opportunities \( O_j \) with a discounting function \( g \), an
example is given below, where the discounting factor \((\beta_1)\) is negative, and the scaling factor \(\beta_0\) is positive:

\[
g(O_j) = \beta_0 \exp(\beta_1 O_j)
\]

(32)

This relates to the intervening opportunities model of trip distribution. (Cui and Levinson 2019b; Stouffer 1960)

### 6.4 Discounting Opportunities due to Competition

When we say there are a certain number of stores available, those stores are assumed to be available independent of the number of people who might avail themselves of those stores. While there might be crowding, it is uncommon and typically neglected. In contrast, when we say there are a certain number of jobs available, those jobs are being competed for by other workers. So seeking a job in a place with say 1,000,000 jobs within 30 minutes and 1,000,000 workers is very different than seeking a job in a place with only 500,000 workers, or one with 2,000,000 workers. Those other workers diminish the available jobs.

Following on the pioneering work of Shen (1998), there is a growing literature on considering competition in access measures. The destination-based access ratio is formulated as the ratio of opportunities reachable from a location, to the demand reaching the same opportunities – the passive access from that location, as described in subsection 2.2. The destination-based access ratio differs from the origin-based measure, in accounting for the source of demand; opportunities at each destination are discounted by the destination’s corresponding demand. The basic formulation of demand adjusted potential access is shown in Equation 35. The formulation is essentially a summation of ‘supply to demand ratio’ for each destination, subjected to an impedance function. Refactoring the presentation of Shen (1998) to be consistent with ours:

\[
A_{i,h,j\text{,workers},m,e,t,p}^O = \frac{\sum_{k=1}^{K} O_{k,h,\text{workers}} f(C_{k,j,h,m,e})}{\sum_{k=1}^{K} O_{k,h,\text{workers}} f(C_{k,j,h,m,e})}
\]

(33)

We define the discounted number of opportunities due to competition as:

\[
g(O_{j,h,\text{jobs}}) = \frac{O_{j,h,\text{jobs}}}{A_{i,h,j\text{,workers},m,e,t,p}^O}
\]

(34)

This results in access to jobs considering competition as:

\[
A_{i,h,j\text{,jobs},m,e,t,p} = \sum_{j=1}^{f} g(O_{j,h,\text{jobs}}) f(C_{i,h,m,e})
\]

(35)

It should be noted that discounting jobs for competition means the number of discounted jobs reachable is on the order of one, rather than perhaps hundreds of thousands, as the number of competing workers is the same order as the number of available jobs. This is of a very different scale than discounting, say, retail opportunities because the first store is more valuable than the second. This is important when trying to construct a multi-activity accessibility measure, as the opportunities are not naturally of commensurate scales.

A number of other methods can be used to account for competition, including the balancing factors from the spatial interaction model (Geurs et al. 2003; Wilson 1971), competition factor (Van Wee et al. 2001), and the two-step floating catchment area method (Luo and Wang 2003).
Towards a general theory of access

<table>
<thead>
<tr>
<th>Mode</th>
<th>Transit Access</th>
<th>Auto Access</th>
<th>Transit Share</th>
<th>Multimodal Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>10,000</td>
<td>100,000</td>
<td>9.1%</td>
<td>91,818</td>
</tr>
<tr>
<td>After</td>
<td>20,000</td>
<td>100,000</td>
<td>16.7%</td>
<td>86,667</td>
</tr>
</tbody>
</table>

Table 2: Illustration of Multimodal Accessibility Paradox

7 Mode

The dimension of how we travel (or by what mode) \(m\) indexes access by whether the trip is made on foot, by bike, by public transport, or by automobile (or by any number of numerous other modes we might consider). Clearly speeds vary by mode. The perceived speed differs from the objectively measured speed. Further, people have different modes available, and perceive the quality of the trip differently.

Accessibility differs by the mode of travel. Sometimes the access of a location needs to be measured while accounting for different modes. Accessibility by different modes can be measured and analyzed separately as we discuss above. Here we look at how to consider them collectively.

Combining the modes is tricky. While activity and time-of-day affect opportunities available, and time-of-day also affects travel cost, mode only affects travel cost and not opportunities. The modes cannot simply be summed, otherwise introducing a new mode would increase access, even if it did not improve service. This is referred to as the ‘Red Bus, Blue Bus Paradox’ in the mode choice literature. (Ben-Akiva and Bierlaire 1999)

Decay functions depend on both mode of travel and trip purpose (Iacono et al. 2008b), and often vary in the speed of travel, affecting access measure. Depending on location, the mode of transport may have a more significant effect on access than the physical distance to destinations (Wachs and Kumagai 1973).

7.1 Weighting Modes by Shares (and a paradox)

Accessibility can be expressed as a mode-share weighted average \(X_{ij,m}\) by different modes; (Knox 1978) as shown in Equation 36:

\[
A_{i,j,h,z,e,t,p}^X = \sum_{j=1}^{I} \sum_{m=1}^{M} X_{ij,m} O_{j,h,z} f(C_{ij,m,e})
\]  
(36)

where the mode share \(X_{ij,m}\) of mode \(m\) for origin-destination pair \(ij\) is constrained:

\[
\sum_{m=1}^{M} X_{ij,m} = 1
\]  
(37)

It has been claimed that aggregating automobile and transit access weighted by mode share provides a reasonably robust measure of access for a neighborhood, that accounts for automobile ownership. (Knox 1978)

The risk with Equation 36 is that the mode share for a slower mode might rise, lowering the overall value of access. In the example in Table 2, there are two modes, and the transit improvement, which doesn’t affect auto access, by attracting additional users would reduce multimodal access from 91,818 to 86,667. This is a version of Simpson’s Paradox.

However, the potential drop in overall access by improving slower modes doesn’t necessarily mean the users are worse off; since the slower modes are often associated with lower monetary costs, the free choice by the users to switch to slower modes reflects their perceived higher utility with the slower modes. It may also come with lower social costs (subsection 3.4).
7.2 Mode with Maximum Accessibility (Lowest Cost)

Assuming all modes are available, access with multiple modes can be expressed by the mode that provides maximum access \(A^*_{i,b,z,,e,t,p}\) in a particular market, shown by Equation 38:

\[
A^*_{i,b,z,,e,t,p} = \max_m \sum_j O_{j,b,z} f(C_{ij,m,e})
\] (38)

Evidence indicates the higher the access a mode provides relative to other modes, the more people tend to choose it (Owen and Levinson 2015). But that is far from saying that everyone chooses the mode with maximum access. Using just personal costs, the mode with highest access to jobs turns out to be auto for almost every Census block in the United States, according to the results of the various Access Across America studies, as the travel time by auto is lower than other modes for almost every origin-destination pair (Murphy and Owen 2019; Owen and Levinson 2014; Owen et al. 2014, 2018). Yet while auto mode share is obviously high in the United States, it is not 100%. This is because people don’t just minimize travel time, and on individual trips people are concerned about the factors of the trip to the particular destination, not access to all destinations. It is also because not everyone has every mode available.

Full cost access (subsection 3.4) offers another way through. With full cost access, considering the full cost of travel \(C_{ij,b,m,e}\), including both short run and long run private and social cost elements \(e\), modes other than the automobile begin to be competitive, as the automobile has greater externalities than other modes. (Cui and Levinson 2018, 2019a) Although full cost access may give a useful value for normative evaluations (e.g., what investments a public agency should make), it doesn’t explain behavior well, as travelers don’t actually incur external costs.

7.3 Modal Availability

The maximum access approach assumes all modes are available. People without automobiles have a reduced choice set compared to those with automobiles. People who cannot ride bikes similarly have fewer options. In such cases an availability matrix is employed to quantify availability-weighted mode-optimal access \(A^*_{i,b,z,,e,t,p}\). With the availability matrix, Equation 39 measures the maximum access of a particular location, from the modes that are available to this individual.

\[
A^*_{i,b,z,,e,t,p} = \sum_{m=1}^{M} \sum_{j=1}^{J} V_{i,m} O_{j,b,z} f(C_{ij,m,e})
\] (39)

The value for each \(i, m\) element in the matrix of modal availability \(V_{i,m}\) takes a value between 0 and 1 indicating the share of people at origin \(i\) for whom mode \(m\) is available as their best option.

\[
V_{i,m} \in (0, 1), \quad \sum_{m=1}^{M} V_{i,m} = 1
\] (40)

\[
V_{i,auto} = \text{share of people with auto available}
\]

\[
V_{i,\overline{auto}} = \text{share with auto unavailable}(= 1 - V_{i,auto})
\] (41)

We first face the question of what it means to “have a mode available”. A wealthy person may forgo car ownership but retains the option to own one may be thought to have long run availability but not short run. In a world with shared vehicles on-demand, the meaning of modal availability also changes.

We then face the question of identifying the “best option,” which presumes having already solved this problem for all modes. As noted, from an individual perspective, in the United States, auto almost always produces higher accessibility (has lower per trip costs) than public transport, so we can think of this as an auto availability measure. In practice, we may make the simplifying assumption that all
other users take public transit or bicycle as the next best alternative. We might write the costs that are used in the impedance function as:

\[
C_{ij,\text{auto},e} = \min_m C_{ij,m,e} \\
C_{ij,\text{auto},e} = \min_{m \neq \text{auto}} C_{ij,m,e}
\] (42)

This would get more complicated, beyond the scope of this paper, if a full cost accessibility approach were used, as auto users incur the internal (private) costs, and behave based on those costs, but produce external (social) costs which are higher; but not all people who have auto available are auto users.

This method avoids the paradox (subsection 7.1) of improving service increasing the mode share of transit and thus decreasing accessibility because even if mode share of transit increases, modal availability of automobile does not necessarily decrease. However it may happen that people give up their cars or the income mix in a neighborhood changes. But in that case the loss of accessibility is real, at least in the short run, not paradoxical.

8 For Whom

The question of access for whom \( (p) \) relates to where (section 2). At the limit, looking at individual persons and highly localized places \( (i) \) will be identical, at least at the points in time where the individual occupies that place. Keep in mind that aggregating spatially (where the subgroup is people who are adjacent to each other at a point in time, e.g., residents of a block) is a very special case of aggregating by groups generally. In the case of equity, we might be interested in places, but we have interest in other kinds of groups. We may be interested not in individuals as such, but in subgroups, for instance minorities, or disadvantaged people, and ask how their access compares with other subgroups, or the population at large. (Martens 2016; Palmateer 2018; Yeganeh et al. 2018)

Most access applications use access to jobs, and this is often taken as a surrogate for general access. On the one hand, access to jobs is relevant not just for work trips, since the place where someone works is often where someone else receives services, so transit access to jobs reflects a general measure of “urban opportunities” (Merlin 2017), and more is preferable to less. On the other hand, we recognize that different job categories represent different economic activities, and different types of jobs generally require different skills and education levels (Geurs et al. 2003; Giuliano et al. 2010). The balance between the number of jobs in each category and the number of eligible workers affects access measures. (Wachs and Kumagai 1973) We compute location-based access by jobs category \( (A_{i,\text{jobs}_k,m,e,t,k}) \) by computing the number of jobs in a particular category \( (\text{jobs}_k) \) reachable within the given time threshold.

\[
A_{i,\text{jobs}_k,m,e,t,k} = \sum_{j=1}^{l} O_{j,\text{jobs}_k} f(C_{ij,h,m,e})
\] (43)

We develop a person-weighted average to all jobs considering the number of workers in category \( k \) at \( i \).

\[
A_{i,\text{jobs}_k,m,e,t,v} = \frac{\sum_{k=1}^{K} S_{i,k} A_{i,\text{jobs}_k,m,e,t,k}}{\sum_{k=1}^{K} S_{i,k}}
\] (44)
Similarly, we can develop person-weighted averages of access to jobs in a category at a location \((i)\) that members of a subgroup \((k)\) can avail themselves, for the entire region.

\[
A_{i,h,job_{i,k},m,e,t,k} = \frac{\sum_{i=1}^{I} S_{i,k} A_{i,h,job_{i,k},m,e,t,k}}{\sum_{i=1}^{I} S_{i,k}} \quad (45)
\]

If the subgroups vary in proportion by location (groups cluster spatially), then the person-weighted average for each subgroup will vary.

### 9 Generalizing Access

In this section we make explicit a combined expression of regional, person-weighted access for multiple time slices, multiple activities, multiple modes, and multiple groups, considering specific cost elements \((e)\) and for a specific impedance function threshold \((t)\). This consolidates the extensions of primal access developed in sections 5, 6, 7, and 8. The resulting expression follows:

\[
A^{\ast}_{i,z,\ldots,e,t,\ldots} = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} \sum_{z=1}^{Z} \sum_{l=1}^{H} \sum_{p=1}^{P} \sum_{e=1}^{E} S_{i,p} U_{h} W_{z} V_{i,m} g(O_{j,z,h}) f(C_{i,j,h,m,e,t})}{\sum_{p=1}^{P} S_{i,p}} \quad (46)
\]

Obviously this explicit expression is cumbersome. If we think in terms of matrices rather than vectors, we can represent the above as a measure of general access:

\[
A_{i} = \sum_{j=1}^{J} g(O_{j}) f(C_{ij}) \quad (47)
\]

In one sense, Equation 47 more or less repeats the original Hansen equation (Equation 1), just making everything bold. We do transform Opportunities \((O)\) by function \(g\) to reflect diminishing value of opportunities and competition.

But what is implied by that is that we are dealing with the matrix of weighted opportunities \((O_{j})\), considering all of the different activity types \((z)\), by all times-of-day \((h)\), and all available modes \((m)\), as well as the matrix of all of the relevant cost elements \((C_{ij})\), which may be the full social costs or generalized internal costs, depending on the purpose.

This is an important shift in thinking from partial to general access. However, we recognize that apart from likely being more expensive in terms of compute costs and the extensive and higher resolution disaggregate data required, the ensuing privacy issues pose challenges to the implementation of a general access measure.

### 10 Lifecycle Access

Access does not simply occur at a single point in time, it is effectively a continuous flow of services provided over many years. When making long-term decisions like investments, people should consider not just today’s access, and not just the access in \(L\) years, but the stream of access over time. We denote
Towards a general theory of access

this as $A_{t, A}$. In this case, the measure might be 'discounted opportunity-years' that a project provides, where $r$ is the discount rate, $l$ is an index of years.

$$A_{t, A} = \sum_{l=1}^{L} \left( \left( \sum_{j=1}^{J} O_{i, j} f(C_{i, j, l}) \right) \Lambda(r, l) \right)$$

A typical discounting function ($\Lambda$) is:

$$\Lambda(r, l) = (1 + r)^{-l}$$

Similar to how the present value of the capital is projected into the future, the opportunity cost from experiencing low levels of accessibility can accumulate over the years. The benefits (costs) of high (low) levels of accessibility experienced early in life may accumulate, so the rich get richer (better opportunities open doors to even better chances) and the poor fall farther behind. Alternatively, those benefits may dilute over time. The lifecycle access provides a pathway to account for the effect of access over time.

Consider for instance two alternative investments. The first provides accessibility benefits now, and for 30 years, the second is capital intensive and only provides benefits in years 25-50. Conventional cost-benefit or cost-effectiveness analysis would account for these time differences through discounting, but benefits might be accounted for as travel time savings. To date, we are unaware of any accessibility analysis that has ever done this, and so accessibility cannot be properly accounted for in project evaluation.

Other appropriate ways to discount the benefits and costs of the future back to the present could be considered, such as hyperbolic discounting, depending on the time value of money, uncertainty, and other factors.

11 Conclusions

People and organizations will pay a premium to be in locations with higher access to the things (people, opportunities) they care about, to save time and money on travel, and to be more productive and earn more, all else equal. The general theory of access makes the claim that access, properly and completely measured, explains location-related decisions.

In theory, we can construct a measure of access to each thing (activity, person, place), by time-of-day, by every available mode, considering every (perceived) cost, for each person considering their unique preferences, at every moment. This looks a lot like what utility theory-based models try to do.

In practice, we construct a measure of access to jobs, (and a few other things) in the peak, by auto and transit (and walk and bike), considering actual travel times (and maybe financial outlays), for the average person or selected groups.

To the extent the practical measure correlates reasonably well with the theoretical ideal, the distinction is less critical. But it is the job of the academics to understand the world better, while others need to put in place something that is actually feasible if known to be imperfect. We delved in this paper to understand what an ideal measure might consider. To reduce the gap between the theory and practice in accessibility measurements, more data covering both a longer time frame, and more accurate demographic and land use data would be necessary, for starters. Whether that additional information required by the more accurate ideal measure is worth the effort of obtaining and analyzing the required data inputs is an empirical and economic question beyond the scope of this work.

Undoubtedly, future researchers will disentangle new dimensions of access.

Already, the concept of virtual access (access via the internet, delivered services, and the like) has received some attention, but doesn’t yet neatly fit in this framework. Access to things includes both going and getting things (going to the store), and having things brought to you like water, natural gas,
electricity, internet bandwidth, and delivered goods, or having things taken away like sewer and waste collection (or, reframing, bringing the absence of filth). These delivered goods, now more prevalent with the internet, are insufficiently recognised as an important aspect of the value of places. Access to input goods and distribution through a supply chain also has relevance at the level of the firm.

That people are more productive in places with higher access to other people, places commonly referred to as cities, implies that access produces externalities that are not fully captured by the real estate markets. We mention a few additional positive externality processes associated with access that may not be fully captured economically. High transit service attracts more ridership, which further reinforces service provision, and existing transit ridership (Bar-Yosef et al. 2013; Levinson and Krizek 2018; Mohring 1972). Accessibility by active modes and to street-fronting retail adds ‘eyes on the street’, which improves the safety of neighborhoods, making communities more livable (Jacobs 1961).

The full internal and external benefits of access are not yet well understood and are difficult to quantify. Accessibility to different types of destinations, by different population groups, and at different time-of-the-day would certainly produce varying levels of spillover benefits. Here we only mention full benefits with a brief discussion to raise the awareness of the topic.

Also undoubtedly, academics will complain about the imperfections of practice, just as the practitioners complain about academics. We cannot let the perfect be the enemy of the good, nor can we let the good stop us from trying to discover the perfect.

We hope this paper generates new directions for accessibility research and analysis, raises new questions for discussion, and places accessibility on a sounder theoretical foundation.

12 Formal Definitions

When formally talking about access, it is critical to have a unifying language with which to speak, where each symbol means one and only one thing, and each thing is defined by one and only one symbol. The language we use throughout is summarized in Table 3 for convenience. There are many equations presented here, but they are far from a complete set. However, we hope the language for talking about access we introduce provides a vocabulary to derive and describe additional measures in a systematic way. Very often nomenclature has been changed from sources from which various definitions were written in the original paper in order to be internally consistent within this document, and avoid repetition of variable names and the concomitant confusion.

The ‘.’ (dot) subscript notation indicates that the dimension is weighted and summed, so the measure applies for that entire dimension, rather than aspects of that dimension.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>§</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0, \beta_1$</td>
<td>Scaling and discounting coefficients.</td>
<td>6.3</td>
</tr>
<tr>
<td>$\eta_e$</td>
<td>Perceived weight for travel time cost elements. $e$</td>
<td>3.2</td>
</tr>
<tr>
<td>$\theta_m$</td>
<td>Impedance factor for mode $m$.</td>
<td>4.3</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>Lifecycle discounting function.</td>
<td>10</td>
</tr>
<tr>
<td>$A_{i,h,z,m,e,t,p}$</td>
<td>Accessibility at location $i$, time $h$, activity $z$, by mode $m$, considering cost elements $e$ for group $p$. $\sum_j$.</td>
<td>2.1</td>
</tr>
<tr>
<td>$A^X_{i,h,z,e,t,p}$</td>
<td>Mode-share weighted primal access.</td>
<td>7.1</td>
</tr>
<tr>
<td>$A^*_t$</td>
<td>Multiple time slice, multiple activity, maximal mode access for system.</td>
<td>7.2</td>
</tr>
<tr>
<td>$A^{**}_{t,v,v,e,t,p}$</td>
<td>Multiple time slice, multiple activity, maximal mode access (primal) for system, considering mode availability ($V_{i,m}$).</td>
<td>7.3</td>
</tr>
<tr>
<td>$A_i$</td>
<td>Matrix of full accessibilities at $i$.</td>
<td>9</td>
</tr>
<tr>
<td>$A_{i,A}$</td>
<td>Accessibility over the lifecycle.</td>
<td>10</td>
</tr>
<tr>
<td>$A^C_i$</td>
<td>Passive access (ease of being reached).</td>
<td>2.2</td>
</tr>
<tr>
<td>$A_{i,h,j,w,m,e,t,p}$</td>
<td>Access to jobs discounting for competition from other workers.</td>
<td>6.4</td>
</tr>
<tr>
<td>$C_{ij,m,e}$</td>
<td>Cost of travel between $i$ and $j$ by mode $m$ considering cost components $e$.</td>
<td>3.4</td>
</tr>
<tr>
<td>$C_{i,j,m}$</td>
<td>Matrix of full costs for mode $m$.</td>
<td>3.4</td>
</tr>
<tr>
<td>$D_z$</td>
<td>Duration of activity $z$.</td>
<td>6.2</td>
</tr>
<tr>
<td>$e,E$</td>
<td>Index of cost elements, Total number of cost elements.</td>
<td>3.4</td>
</tr>
<tr>
<td>$e_t,E_t$</td>
<td>Index of travel time elements, Total number of travel time elements.</td>
<td>3.1</td>
</tr>
<tr>
<td>$e_{int},E_{int}$</td>
<td>Index of internal cost elements, Total number of internal cost elements.</td>
<td>3.3</td>
</tr>
<tr>
<td>$e_{full},E_{full}$</td>
<td>Index of full cost elements, Total number of full cost elements.</td>
<td>3.4</td>
</tr>
<tr>
<td>$f(C_{ij,m,e})$</td>
<td>Impedance function.</td>
<td>2.1</td>
</tr>
<tr>
<td>$g(O_j)$</td>
<td>Opportunities discounting function.</td>
<td>6.3</td>
</tr>
<tr>
<td>$h,H$</td>
<td>Index of times-of-day (time slices), total number of time slices.</td>
<td>5</td>
</tr>
<tr>
<td>$i,I$</td>
<td>Index for origins, total number of origins.</td>
<td>2</td>
</tr>
<tr>
<td>$j,J$</td>
<td>Index for destinations, total number of destinations.</td>
<td>2</td>
</tr>
<tr>
<td>$k,K$</td>
<td>Index for category, Total number of categories.</td>
<td>8</td>
</tr>
<tr>
<td>$l,L$</td>
<td>Index for lifecycle (years), Total number of years.</td>
<td>10</td>
</tr>
<tr>
<td>$m,M$</td>
<td>Index of modes, total number of modes.</td>
<td>7</td>
</tr>
<tr>
<td>$N$</td>
<td>Threshold number of activities.</td>
<td>2.1</td>
</tr>
<tr>
<td>$o$</td>
<td>Index of origins.</td>
<td></td>
</tr>
<tr>
<td>$O_{j,z,b}$</td>
<td>Opportunities at $j$ of type $z$ at time $b$.</td>
<td>2.1</td>
</tr>
<tr>
<td>$O_i$</td>
<td>Matrix of full opportunities.</td>
<td>9</td>
</tr>
<tr>
<td>$p,P$</td>
<td>Index for population subgroup, total number of subgroups.</td>
<td>2.4</td>
</tr>
<tr>
<td>$r$</td>
<td>Interest rate.</td>
<td>10</td>
</tr>
<tr>
<td>$s_{ij}$</td>
<td>Distance.</td>
<td>3.1</td>
</tr>
<tr>
<td>$S_{i,p}$</td>
<td>Population at origin $i$ in group $p$.</td>
<td>2.4</td>
</tr>
<tr>
<td>$t_{ab},t_{lb}$</td>
<td>Travel time threshold, upper bound, lower bound.</td>
<td>4.1</td>
</tr>
<tr>
<td>$t_{ij,e,m}$</td>
<td>Travel time between $i$ and $j$ considering time element $e_t$ and mode $m$</td>
<td>3</td>
</tr>
<tr>
<td>$t_{ij,m}$</td>
<td>Perceived travel time.</td>
<td>3.2</td>
</tr>
<tr>
<td>$t_p$</td>
<td>Travel time for person $p$.</td>
<td>3</td>
</tr>
<tr>
<td>$U_h$</td>
<td>Weight of each time slice $h$.</td>
<td>5</td>
</tr>
<tr>
<td>$v_{ij}$</td>
<td>Velocity (speed).</td>
<td>3.1</td>
</tr>
</tbody>
</table>
### Continuation of Table 3

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{i,m}$</td>
<td>Matrix of modal availability at origin $i$ for mode $m$ (0, 1).</td>
<td>7.3</td>
</tr>
<tr>
<td>$w$</td>
<td>Minimum wage rate or measure of value of time.</td>
<td>3.3</td>
</tr>
<tr>
<td>$W_z$</td>
<td>Weight for activity type $z$.</td>
<td>6.1</td>
</tr>
<tr>
<td>$X_{ij,m}$</td>
<td>Share of a given mode $m$ (0,1) in a given market $(ij)$.</td>
<td>7.1</td>
</tr>
<tr>
<td>$Y$</td>
<td>Matrix of activity substitutability, complementarity.</td>
<td>6.1</td>
</tr>
<tr>
<td>$z,Z$</td>
<td>Index of activities, total number of types of activities.</td>
<td>6.1</td>
</tr>
<tr>
<td>$$</td>
<td>Monetary cost of travel.</td>
<td>3.3</td>
</tr>
</tbody>
</table>

**End of Table**
References


Towards a general theory of access


Towards a general theory of access


