The Concept of Rationality for a City

Kenny Easwaran

June 20, 2018

The central aim of this paper is to argue that there is a meaningful sense in which a city can be thought of as a collective agent, to which some concept of rationality can apply. The idea will be that a city is rational to the extent that the collective practices of its people enable diverse inhabitants to simultaneously live the kinds of life they are each trying to live. This has significant implications for the varieties of social practices (including social customs, physical infrastructure, and laws) that constitute being more or less rational. Some of these implications may be welcome to a theorist that wants to identify collective rationality with a notion of justice, while others are unwelcome. There are some significant challenges to this use of the concept of rationality, but I claim that these challenges at the city level have parallels at the individual level, and may thus help deepen our understanding of rationality at all levels.

One important clarification should be made first. The word "city" is somewhat ambiguous. It sometimes refers to a specific legal jurisdiction that is incorporated as a municipal entity. In this sense, Los Angeles County has 88 cities, and the current municipality of Princeton, NJ, was until 2013 composed of two separate cities — the Borough of Princeton (which contained most of the university, and a small surrounding area) and Princeton Township (which completely surrounded the other municipality). The word "city" is also sometimes used (particularly in Australian English, but the word has been borrowed into German and other languages with this meaning) to refer to a small but highly built up central business district of a larger metropolitan area. Century City in Los Angeles is an example of a location referred to in this way.

However, my usage of the word "city" corresponds more closely to the US Census Bureau's notion of "urbanized area". As I will explain later, the important thing I want to track is a community of people whose daily lives are tied together by geography, rather than a governmental entity or a legal border. Importantly, this includes the suburbs and outlying neighborhoods, and not just the very central core. Furthermore, it is essential to my project that the focus is on the collection of *people* defined by their geographic proximity, rather than just the geography itself. The structures and infrastructure built by people (both physical and legal), and perhaps to some extent the physical geography of the space as well, are included in what I term a "city" only to the extent that they play a role in structuring and being structured by the activities of the people.

1   Intuitive motivation

In traveling to different cities, one is struck by some differences in how they work. Some of these differences involve the physical or legal infrastructure of the city, while others involve the social norms exemplified by people present in the city.

Physical infrastructure is perhaps the most obvious. Manhattan is mostly covered with a convenient grid of streets and avenues that aids both newcomers and long-time residents in navigation, and also aids in the development and redevelopment of rectangular buildings that make efficient use of the plots of land they are on. Chicago and Salt Lake City take this to an even greater extreme, while Boston and London do not, to say the least. Less obvious to most people, some older North American cities, like New York and San Francisco, have a single set of pipes that carries rainwater falling on the streets as well as sewage draining from buildings, which can sometimes result in untreated sewage spilling into rivers or even streets during heavy rains. Other cities, like Los Angeles and Dallas, have completely separate sanitary sewers and storm drains. People who want to get where they're going, or who have a well-functioning sense of smell, may say one of these systems is more rational than the other.

Differences in social norms are also quite noticeable. In many cities, like Washington, there is a very strong social pressure to stand still only on the right side of escalators in the Metro system, so the left side can be used for people walking. But in other cities, like Los Angeles, there is no such custom, and standing people often make it difficult to race down the escalator to catch a train. Drivers in Los Angeles will treat a left turn signal as extending for several seconds after it officially turns red, squeezing a few extra turners after the light cycle ends. Meanwhile, in Pittsburgh, even at traffic lights without a protected left turn phase, people going forwards through the intersection will allow a left turner to go first at the beginning of the green.[1] Again, some people might find one practice more rational than the other.

The legal infrastructure of a city contributes to shaping both the physical infrastructure and the social norms. The streets and sewers of most cities were built by the legal and governmental structures that exist there. Practices of driving on the right or left, crossing streets at marked intersections, and so on, are usually heavily shaped by local laws. However, many social practices (like the different left turn patterns in Los Angeles and Pittsburgh) exist in contravention of local laws. And even ones that fit the laws often cross municipal,

---

[1] Although I have been unable to find empirical studies on the prevalence of these behaviors, the Los Angeles left has been described as "part of driving culture" in the Los Angeles Times (Bernstein, 2003), and a writer for LA Weekly says "many Angelenos, this reporter included, were raised to believe in the so-called "third car rule," which states that the third car can turn left on a red pretty much no matter what, unless the driver of the second car falls asleep or something." (Aron, 2017) The "Pittsburgh left" even has its own wikipedia page. Online discussions (for example: `https://ask.metafilter.com/15392/Left-turns-in-Los-Angeles`) have suggested that each of these practices occurs in other cities as well, but they appear to be most commonly attributed to these two cities.

I thank an anonymous referee for pressing me to find citations for these practices.

county, and even state and national borders.

Some philosophers claim that a notion of "rational" that can apply to cities is merely metonymy.

> In ordinary speech we evaluate a vast array of different things as rational or irrational: people, dispositions, habits, emotions, and even laws, city layouts, voting systems, arguments, and conversations. . . . The hope will be that insofar as we can talk about the rationality of laws, emotions, dispositions, city plans, arguments, and the like, this will be derivative on the rationality of beliefs, desires, and actions, as things that either result from or cause irrational beliefs, desires, or actions (though perhaps city plans and laws are only irrational insofar as they are inefficient or arbitrary). (Hedden, 2015, p. 10)

However, my claim is that there is a kind of literal truth here that is worth investigating. Furthermore, I will claim that the relevant notion of rationality extends to far more than the obvious features mentioned in these examples.

2 Individual and group conceptions of rationality

2.1 Individual rationality

The particular conception of rationality I work with is a sort of means-ends practical rationality. On my notion, someone who has a given set of desires is considered rational to the extent that their behaviors are well-governed by their desires. (I don't particularly distinguish desires, goals, or other similar end-type states.) To be an agent is to be the kind of being whose behaviors are governed by desires, and to be rational is to be a good agent. Aspects of this notion of rationality have been considered important by many philosophers.

Some philosophers have also argued for further aspects of rationality governing what sorts of desires a rational agent could have. However, I will focus on a more minimal notion of rationality that is more akin to that found in Hume — any desires whatsoever could be those of an agent, and the agent is rational to the extent that her behaviors are well-governed towards realizing those desires, regardless of how bad we might think those desires are. Famously, Hume argued that there are no substantive constraints of rationality on desire — there is no need to desire one's own survival or one's own future pleasure, or the happiness of others.

Epistemologists also discuss rationality, and there are parallel debates there between "permissivists", who argue that agents with different "prior beliefs", or "epistemic standards", can rationally believe different things on receiving the same evidence, and supporters of "uniqueness", who argue that there is one uniquely right response to any body of evidence. (Kopec and Titelbaum, 2016, Horowitz and Dogramaci, 2016, Schoenfield, 2014) Since I focus here on

collections of people as agents, and epistemic notions aren't totally clear in this collective setting, I will try to avoid issues relating to the interaction of these types of rationality. (Briefly, one might think that an agent would be rational to do A, if she had a rational, but false, belief that A is the behavior most likely to promote her desires, even if B is the behavior that would *in fact* promote her desires best.) However, much of what I say will naturally go along with a permissivist epistemology, in line with my Humean view on desires.

However, the Humean and permissivist aspects of my account are somewhat tempered, for reasons that I will explain in somewhat greater depth in section 3.2. Although agents might have any intrinsic desires, there will be some instrumental desires that tend to be widely shared. For instance, although I say some agents might lack an intrinsic desire for their own survival, they are likely to recognize that at least some of their desires are idiosyncratic (for instance, a desire for their own children to thrive, or a desire that a particular project be completed) and these desires will be better promoted by their own future behavior than by that of random other people. Thus, in most contexts, even someone who lacks an intrinsic desire for survival will usually have an instrumental desire to survive, at least in order to help carry out these projects. As I will argue later, there is a lot more that tends to be shared by human beings living in a shared geography, as a city.

2.2   Collective rationality

Given this means-ends conception of rationality, for some entity to have the capacity for rationality, it must have desires, and it must have behaviors that are motivated by those desires in the right sort of way to be evaluated for rationality. Thus, for there to be a meaningful sense in which a group can be said to behave rationally or irrationally, there must be group desires, and group behaviors that could be well-governed by those desires. What exactly these two conditions amount to in my usage will take some explanation. However, I don't believe that there is only one notion of group rationality, for which we must find *the* correct specifications of "group desire", "group behavior", and the way the one shapes the other. Rather, I think any such specification will give *some* notion of group agency and group rationality, and the question is whether some particular specification gives rise to a notion that is interesting enough to be worthy of study. I claim that applicability to cities does enough to motivate the versions that I consider, and may help clarify the particular account of group desire and group behavior I have in mind.

The version of "group desire" I care about is a relatively minimal one. It is sufficient for the existence of a "group desire" that the members of the group happen to have the same desire. There is no need for the individuals to desire it in some special group-oriented way. There is no need for each individual to even be aware that the others share the desire. There is no need for them to desire that the others achieve their desire.[2]

---

[2]While some conceptions of collective attitude allow for a group desire to exist even without the members of the group sharing it, they usually require some greater degree of structure or

The paradigmatic desires I will concern myself with here are indexical and instrumental. These are desires like, "that I can get across town quickly when I want to", and "that I can have air to breathe that is low in particulate matter", and "that I can easily get clean drinking water". These desires are indexical, because while considerate people do have these desires for many others, they are usually strongest for themselves (and their closest loved ones), and many people actively desire the opposite of these things for particular other individuals they dislike. These indexical desires are shared in the sense that each person will have a parallel indexical version, and not in the sense that people desire the abilities for the same person.

I say that these desires are instrumental because they are subservient to other individual ends. While there are some people who have an intrinsic desire to be in multiple distant places at different times, for most people, the desire to be able to cross town quickly and conveniently is instrumental. One person might want it because his mother lives across town and he wants to see her. Another person might want it because she heard there's a new Thai restaurant across town and she wants to try it out. Yet another person might want it because this crosstown trip is their commute between home and work. Each of these people would be just as happy if the object of their desire came to them, or happened to be located in geographic proximity to them, but being able to travel easily across town is a good way to achieve this desire and others.

The desires for clean water and air might seem more fundamental, but I claim that for most people, they are also merely instrumental. I will discuss this further in section 3.2, but the basic idea is that clean water and clean air are valuable primarily because the lack of them causes health problems that stop people from achieving their other desires. (Many of us also find clean air and clean water to be pleasant, but there are others who intrinsically prefer the smell of cigarettes or fireplaces or industry, and the taste of alcohol or soda.)

The version of "group behavior" I care about is a little thicker, but not much. What seems essential to me for evaluations of practical rationality is that there be some way that the behaviors are responsive to the desires. Thus,

---

self-conception for the group than I do. (For instance, Searle (1990), Gilbert (1990), and other accounts described as "non-summative" by Tollefsen (2002).) Thus, I consider my condition more minimal than these others, even if the requirement that the individuals share the desires seems strong from another point of view.

Given the particular desires that I will discuss, about the basic functioning of the infrastructure of everyday life, like air, water, and transportation, it really is plausible that most or all residents of a city share them. But as I will discuss in section 4.2, there may be some sub-communities that are excluded from membership on this account of the group, because they don't share these desires.

Some might worry that even though each particular desire I discuss is common, it might be the case that most residents lack at least one of the whole set that I use to characterize the city. In that case, it might be important to modify my condition for membership to be one in which each individual shares *most* of the desires, or to consider membership in the group as coming in degrees proportional to the extent of shared desire, so that it is plausible that my group corresponds closely to the set of residents of an urban area.

I thank an anonymous referee for suggesting this more detailed discussion of the sense in which my requirement is "minimal".

for a collection of agents to count as a further collective agent that is evaluable for rationality, there must be enough interaction among the agents for the fact of their collective desires to shape their collective behavior. I claim that much about the way cities work exemplifies this idea.

Some of the obvious ways in which the groups of people making up a city can coordinate their behavior in ways that serve their collective desires involve explicit coordination through various levels of government and formal social organizations. However, I think this is only a relatively small amount of the way that coordinated behavior arises in a city. In nearly any dense urban area with many municipalities, the official jurisdictional lines are hard to notice — the most obvious indication that one has crossed between the cities of Los Angeles, West Hollywood, and Beverly Hills is the color of the street signs; the pattern of two cars turning left after the left arrow turns red persists in all three municipalities. There are distinctions that matter if one happens to have a child in public school, or a reason to interact with the police, but not in terms of most of the habits of speech, walking, driving, and otherwise interacting with other people.

Instead, I think that mere physical presence and repeated interaction with a community of interacting people is the way in which most desires get translated into action. Through mechanisms like those described by Muldoon et al. (2013), behavioral regularities can emerge whenever there is some set of shared interests. Many of these ways in which we allow our behavior to be shaped by the behavior of others is not consciously chosen, but it still enables a way for collective desires to shape collective behavior, and thus allows a kind of evaluation of collective agency and rationality.

Margaret Gilbert explicitly rules out the thin kinds of group desire and group behavior that I consider, in the early pages of her "Walking Together". (Gilbert, 1990)[3] She considers a case where two people walking along a path notice each other's presence, and each desires for their mutual presence to continue, perhaps because each has noticed the other around and wants to get to know the other. She says this "weak shared personal goal" is not sufficient for their behavior to constitute an act of walking *together*. She further says that even a "strong shared personal goal", where there is common knowledge that they share the desire, is not sufficient. Instead, she requires that the desires have a plural subject — that *we* continue this walk together, rather than merely that *I* continue to walk in the presence of the other.

Furthermore, this shows up in the way that she thinks the behavior should be responsive to the desire. I merely require that the behaviors of the group be *causally* responsive to the shared desires. For example, this could be achieved for walking together if each individual can make a noise that will have the effect of slowing the other down if he gets ahead, or for the organization of city streets by means of any of many social uses of a car horn. But Gilbert requires instead that each member of the group should be *entitled* to *rebuke* the other

---

[3]As a referee notes, there are many other relevant accounts to contrast mine with beyond that of Gilbert, but I focus on her as one convenient foil. For more detailed discussion of several such accounts, see Tollefsen (2002).

for failing to keep their collective action in line with their collective desire. This is certainly the case when we explicitly go for a walk together (the noise I make when you get ahead might be the verbalization of the rebuke), but I allow for weaker versions that she explicitly rejects.

This particular weaker conception of agency that I describe is less than many philosophers of action have considered necessary for real action. Consider Davidson's example of a rock climber who is climbing with someone that he wants to kill. Awareness of the desire to kill the person may make the climber's hand tremble while he is supporting the rope, causing him to drop the other person to his death. Davidson points out that we wouldn't ordinarily want to count this as a case of intentional behavior, subject to evaluation for rationality. My account does appear to allow for cases with some of this irregular causal structure.

This particular case is one in which the climber exercises little or no active control over the behavior. It might seem that a conception of action and normative evaluation that includes this case is more Freudian than one might like. However, I think there are other cases of relatively low direct control that really are subject to some sort of normative evaluation. Much of what it takes to be a good athlete involves tuning one's reflexes so that they respond instantly to situations where a fast response is necessary, and so that they don't respond to situations where one wants to hold one's ground. The same is true for being a good painter, surgeon, masseur, research scientist, ninja, or parent. Someone who wants to be a microbiologist and *doesn't* do the work of training her hands to be steady when pipetting tiny samples is guilty of *some* sort of irrationality. While there is some sense in which we don't want to say she is irrational in the moment of the trembling hand that causes her to spill her sample, there is also a sense in which this moment is a manifestation of a kind of irrationality in her extended plans.

Returning to the case of groups, Gilbert is interested in a notion of rationality for relatively tightly organized, coherent groups, like teams or corporations. She takes the example of walking together as a model for how these organizations come about. I am interested in a broader notion of rationality that applies to groups with more diffuse structure, maintained mainly by geographic proximity and some basic shared human needs and capacities. It may be helpful to think of the notion I discuss as "skill" or "ability" (or "intelligence", following Bostrom (2012)) rather than "rationality", but these notions have some close relation in any case.

2.3   Analogies to diachronic rationality

Conditions like the ones I take to constitute collective agency have also been drawn in analogy to conditions for diachronic epistemic rationality. Matthew Kopec (2015) draws analogies between a discussion of diachronic Dutch books in David Christensen (1991), and a discussion of interpersonal Dutch books in Donald Gillies (1991). Christensen takes it as axiomatic that there are no epistemic requirements of interpersonal coherence, and finds the similarity to diachronic

Dutch books to raise problems for diachronic requirements of rationality. Gillies, on the other hand, uses the analogy to argue that scientific communities ought to have some sort of intersubjective agreement on a probability function that can then be used to make confirmation theory less subjective. Kopec tries to sharpen up the conditions under which such a community has a requirement of epistemic coherence.

Gillies gives two conditions that bear a great resemblance to the two conditions I give:

> (i) (Common Interest) The members of the group must be linked by a common purpose; whether the common purpose leads to solidarity or rivalry within the group does not matter much; the important point is that the members have an interest in acting together and reaching consensus. ...

> (ii) (Flow of Information) There must be a flow of information between the members though it does not matter whether the communication is organized centrally or peripherally or whether it is direct (between any two members) or indirect (through the intervention of third parties). (Gillies, 1991, p. 518-9)

"Common Interest" is a somewhat stronger version of my shared indexical instrumental desires. "Flow of Information" is a particularly epistemically-directed version of a way for behavior to be governed by these shared desires. Kopec strengthens the first condition and adds some further conditions to determine the scope of the propositions that the group will form probabilities over, and treats the second condition as merely one means to a possible end of coherent belief.

If we consider a collection of entities, whether time-slices of an agent, or separate agents in a group, versions of these two conditions give rise to an argument (known as a "Dutch book" in the literature of formal epistemology) for particular conditions on rationality. We can think of a "credal state" regarding p as a disposition to take actions that satisfy one's desires to a greater degree if p is true than if p is false, and that the "strength" of this credal state is the particular weighting of these two types of outcome when choosing between possible acts. Under this interpretation of credal states, we can argue that an agent, or collection of agents, with a particular set of desires will have dispositions that result in guaranteed poor performance according to these desires, iff the relevant credal states fail to satisfy the axioms of probability. For an individual at one time, her credal state must satisfy the axioms of probability theory, (Easwaran, 2011) and for a group, their individual credal states must be identical. (Kopec, 2015) When there is no way for these dispositions to coordinate with one another in this way, (perhaps because the group lacks communication methods, or because the individual is under the influence of a drug) this is merely a tragedy. But when there is a way for the collection of dispositions to shape each other, the collection is performing poorly to the extent that they don't exhibit probabilistic coherence.

For successive time slices of an ordinary human, there usually is a very large degree of shared concern with the same desires. Furthermore, the operations of memory and intention provide a very strong flow of information. But since this flow of information only works in one direction, the rationality requirement that is most naturally justified doesn't require every time slice to have the *same* degrees of belief, but merely that the later ones are versions of the earlier ones, conditionalized on whatever extra information they have received in the interim. (Teller, 1973, Greaves and Wallace, 2006)

I take this analogy between the interpersonal and diachronic cases to help characterize the cases in which the standard rules fail. Many standard problem cases for conditionalization arise in contexts where the operation of memory or intention is blocked (Talbott, 1991, Elga, 2000, Arntzenius, 2003, Hitchcock, 2004, Bradley and Leitgeb, 2006), or when desires have undergone some sort of transformation. (Paul, 2014)

But some other epistemologists take worries about the real existence of interpersonal requirements of rationality to undermine the real existence of diachronic requirements of rationality as well. Brian Hedden (2015) and Sarah Moss (2015) have argued that we don't need any such argument for diachronic norms of rationality. Instead, only the time-slices of an agent can be evaluated for rationality, and there is no diachronic coherence norm like conditionalization. However, on their view, there are substantive evidential norms regarding what credal states one ought to have given a particular body of evidence, and in the ordinary case, someone who is rational at each time will have a diachronic pattern of credal states that happens to satisfy conditionalization. This is akin to Niko Kolodny's argument in his (2007) that all coherence requirements are just summaries of patterns of substantive requirements of rationality.

However, it seems to me that these sorts of substantive requirements aren't forthcoming, either in epistemology or practical rationality. If the Humean picture is right, that any final ends could constitute the desires of an agent, and if the problems of Carnap's project show that there is no substantive inductive logic saying what credal states are required beyond the ones deductively entailed by one's evidence, then any notion of rationality will at most involve some sort of coherence between preferences, desires, and beliefs. When there are shared desires that can together cause shared action, as in the successive time-slices of an ordinary human, or the members of an urban community, then there is some sort of rational coherence that is required. But when the collection lacks these shared desires, or the means of coordinating behavior, there is merely tragedy, not irrationality.

3 Rationality for a city

3.1 Geography and the human body

I have already mentioned several desires that residents of a city tend to share. They want to be able to get across town conveniently, have clean air, and have

clean water, among other things. And these are things that they want for their own distinctive reasons, whether it be to visit family, go to work, eat some interesting food, or go to the beach. Residents also generally want nearby businesses to stay in business and new ones to open, and they want their neighbors to have stable and law-abiding lives, and they want special event sites like theaters, city hall, and major parks, to remain in existence and to remain accessible.

This is why the characterization I give of the city is the whole urbanized area, rather than the government or the legal boundaries of the municipality. A city is the group of people whose fate is tied up with each other as a result of their geography. Manhattan has 2 million residents. But on an average weekday there are 4 million people there. Even if you never set foot on Manhattan, if you live in Yonkers, or Fort Lee, or Levittown, your fate is tied up in the daily goings on in Manhattan. You will have neighbors or customers, delivery people or doctors, who do go into Manhattan regularly and depend on it for their existence. Conversely, even if you never leave Manhattan, and never go south of 60th St, or never go north of 14th St, your life depends on the millions of people from the surrounding areas who come through every day to do their business, and support the systems that make your life work. Jersey City and Hoboken are as much a part of the New York community as the Rockaways or Staten Island.

Agnosticism about the governmental status is essential not just in order to include the people whose lives cross jurisdictional boundaries, but also to count the people whose presence lacks legal status. Whether it is a Mexican citizen on an expired tourist visa who is working in a factory in North Carolina, or a resident of Beijing with a hukou permit that ties them to a rural part of Gansu province, a person who is geographically present in a city will care that the city functions well, whether or not the city government believes they should be allowed to remain there. These illegally resident people may have different desires from other members of society about the enforcement of some of these legal provisions, but they still want effective transportation, safe and stable local businesses, and clean air and water. Furthermore, many other residents will want success for the businesses that employ these people, and for the stores that they shop at, and will thus indirectly care about their health and well-being, whether or not they care for them in the sympathetic way we hope humans care about each other.

These shared desires are a function of the interaction between human embodiment and geography. If we humans didn't care about touching our loved ones, and eating meals with them,[4] and didn't need physical presence for most of the things we care about, we might not have such a shared concern for efficient travel. Instead, we might be happy with a global network of satellite phones, and each remain in physical isolation on our own isolated farmsteads. If we didn't have to walk to get to the car, and to drive through local streets to get to the freeway, we might not particularly care about the health of local businesses that we pass along the way. If most of the businesses and institutions

---

[4]Kant says that dinner parties are the "highest ethicophysical good". (Cohen, 2008)

we cared about weren't sustained by the activities of strangers, we might only care about effective travel for the specific routes that we personally use. (In many cases we don't explicitly realize we care about these things, but we notice and complain when they fail.)

Furthermore, the way these desires interact to shape behavior is also a function of the interaction between human embodiment and geography. In a crowded public space, the way people do or don't queue up, and the way they do or don't make way for people walking by in the opposite direction, varies greatly from city to city. If humans moved in some way other than walking, and if we didn't need to go through the same places to get to our different destinations, these particular social customs wouldn't be important. (Whyte, 1980)

Many of these issues are extremely shaped by technology. Patterns of human interaction in automobiles are incredibly different from patterns of human interaction on foot or bicycle or in a subway car. Conventions for standing right and walking left don't matter unless there are escalators (so it's no surprise that these conventions are more strongly developed in cities where most people have used underground transportation for decades than in cities where the main experience of an escalator is in a shopping mall).

The physical infrastructure of street and sewer systems that I mentioned at the beginning is a more complicated case. The grid layout of streets in a city, and the construction of pipes of varying sizes and connections underground, are not activities that most residents of an area have any direct involvement with. However, these infrastructural features are usually created and maintained through the various political bureaucracies of the region. Sometimes these are coextensive with the political jurisdiction of one municipality, but in other cases they are not — even in just the urban area of Los Angeles, drinking water for over 300 municipalities comes from one entity (the Metropolitan Water District), while sanitary sewage services run by the City of Los Angeles drain waste from 30 cities, and storm drains are maintained by a combination of Los Angeles County and most of the individual cities within the county.

The street grid of Manhattan was established in 1811 by a commission set up in 1807 by the "Common Council", but was in fact implemented by many more people. Decades of real estate developers, builders, city councils, surveyors, and teams involved in leveling rocks that blocked the paths of streets, were needed to actually carry out the relevant construction. Had the city council in various decades been more lenient towards squatters living in places designated for future parks, and developers that wanted buildings that overlapped planned streets, the effect would have been quite different. As it is, there are several major deviations from the grid in major institutions like Columbia University, Grand Central Terminal, Lincoln Center, and modernist housing projects like Stuyvesant Town. The particular extent of persistence of the grid through many topographic and commercial obstacles is in large part due to the particular culture of civic order that New York had at various points in its history.

The traffic jams around these chokepoints, and the difficulties of moving crosstown at rush hour, are inefficiencies of current behavior that derive from past decisions about infrastructure, rather than from current problems of co-

ordination. These features of a city are analogous to the trembling hands of the microbiologist that never put in the hours in grad school developing skills with the pipette. They may be counted as present irrationalities, or merely as present manifestations of past ones.

San Francisco is not the hilliest city in the United States (in many respects, Seattle and Pittsburgh are hillier, while in others Los Angeles is (Pierce and Kolden, 2015)) but it is famous for its hills because some parts of the central street grid completely ignore their presence and continue straight, regardless of how steep they have to be to do so. The fact that a street grid was established in San Francisco despite these conditions, and that street grids were not established in most late 20th century suburban developments, is in some way constituted by the collective behavior of the residents of those cities in those times. This may be helpful in the ways that grids are often helpful, or hurtful in accentuating the steepness of the slopes. It may be analogous to the person whose habit of early waking sometimes causes pain on weekend mornings after some bumpy nights, even if it serves her well in navigating her daily schedule.

Some of the phenomena I'm talking about exist also at higher and lower levels — one might say that a neighborhood or a nation also has a set of shared interests and some ability to coordinate its actions to bring them about. However, I claim that urbanized areas exhibit a greater level of unity and independence from outside influence than these other levels.

At the neighborhood level, there is often some greater degree of shared interest (residents of San Francisco's Chinatown have a shared interest in the outcome of the Central Subway construction project underneath them, and in the waste management practices of the Ritz-Carlton hotel located on top of a nearby hill, that is of much greater degree than the interest in these issues had by residents of the Sunset or Mission districts). But residents of one neighborhood usually work in another neighborhood, and frequently engage in commerce or entertainment or recreation in other neighborhoods as well. And others who officially live outside the neighborhood often enter it for many of these things, or even just to visit friends in purely residential neighborhoods. These connections tie all the neighborhoods of an urban area tightly together, even across municipal borders and bodies of water. While no individual has daily behavior that travels through the city as a whole, lines tracing out the daily interactions of each individual will very heavily cross the boundaries of any designated sub-region within the urbanized area. (Alexander, 1965) But since most people only leave the urban area for longer trips, there is a natural sense in which the urbanized area is one of the strongest units.

At the national level, there is also some amount of shared behavior motivated by shared interests. However, these tend to be weaker than the ties at the level of the urban area. Most of the conditions of daily life — water, sewage, transportation, air quality, broadcast media, electricity — are supplied through local or regional entities. Much of the legal system, and in some places medical insurance and education, are supplied by state or national level entities. But these tend not to be the conditions of daily life in the same way. And there are many urban areas that manage to spill across national boundaries the same way

cities like New York, St. Louis, and Philadelphia spill across state boundaries. San Diego and Tijuana, and Detroit and Windsor, function as largely unified urban areas. Basel, Switzerland, spills over into both France and Germany. Furthermore, many of the differences in practice that might constitute the rationality or irrationality of a city are concentrated at the level of the urban area, rather than following state or national borders.

3.2   Broadly shared instrumental desires

I have mentioned some instrumental desires that will tend to be shared by humans living in geographic proximity to one another. Some of these relate to transportation (because most things humans care about require spatial proximity to something or someone) and human needs for survival. There are probably also some intrinsic desires that are broadly shared among humans — due to our shared evolutionary history, there are some aesthetic values that we likely all share to some degree or other.

But there are also some instrumental desires that are likely to be shared by *any* agent, whether embodied like a human or not. Some philosophers have motivated the importance of questions about personal identity by assuming that people generally have an intrinsic desire for continued existence, and thus will care whether a particular process results in their own continued existence, or their instant, painless destruction followed by the creation of a qualitative duplicate. Because of the Humean motivational picture I assumed above, I don't think this question will necessarily interest every single agent. However, assuming that an agent is at all effective in bringing about its desires, it is going to put some instrumental value on the continued existence of itself, or at least a qualitative duplicate, since this continued existence is likely to help promote the satisfaction of some of its other desires. (Bostrom, 2012) This won't always be the case — if a parent cares intrinsically about the survival of their child, or a patriot cares intrinsically about the defense of their nation, they may well be willing to sacrifice their own life in order to block a threat to the one they care about.

Stephen Omohundro has explored the question of what instrumental desires would be broadly shared by any agent living in our physical universe, in the context of trying to understand what motivations a superintelligent artificial intelligence would have. (Omohundro, 2008a,b) He argues that any rational agent will have some states we can interpret as beliefs and desires, at least approximately in accord with the standard Bayesian account (or at least, as far as resources allow). Furthermore, any such agent will recognize the fundamental physical importance of certain resources (space, time, energy, and matter), and will try to use them efficiently, prevent the loss of such resources, acquire more of these resources, and seek new creative ways to use these resources to advance its desires. All of this will manifest in some sort of instrumental desires for self-improvement and self-preservation, along with various modifications of the environment to help these processes.

When dealing specifically with humans, the particular resources involved can

be stated more explicitly. Food, water, and air will be useful for humans generally, even if particular artificial intelligences might take matter and energy in other forms to be just as useful. However, the importance of gasoline, electricity, and heating fuel to most households shows some of the way that humans have become more sensitive to energy generally, even in forms that humans don't directly consume.

One other resource that is likely to be of broad use to agents of any sort is information. While Omohundro doesn't explicitly put this on the list of resources, much of "The Basic AI Drives" concerns the way artificial intelligences would try to use information efficiently and acquire more of it while protecting the information it has. For humans, there are more specifics that we can expect about the sorts of epistemic resources we will all tend to find useful — not just any old information is valuable, but information in forms that are accessible to humans is more valuable.

One thing to note however is that if, as I say, there is a sense in which the city is itself to be seen as an agent with desires of its own (i.e., those that are shared by its inhabitants), then the city *itself* will have some of these same instrumental goals. A city, as a community, will have some instrumental desire to protect itself, in order to better serve the shared needs of its residents. This will likely be manifest in the way that each resident of the city cares to some degree that the city remains well-functioning, in order to support her own activities. It will also be manifest in the way that residents of a city usually care about the economic health of the city, and the fiscal health of the local government and its businesses.

4   Specific implications of the view

4.1   Specialized cities

Although cities generally will involve residents who share many instrumental desires, there are some cities that have even greater convergence of desires. This may be because of some shared intrinsic goals of the residents, or because they have shared specialized instrumental goals. As an example of the first sort of case, we might consider a city with homogeneous shared culture, ranging from merely all wanting to observe Thanksgiving Day on the fourth Thursday in November, to something as strong as a monastic community that engages in tightly coordinated religious practice. As an example of the second case, we might consider a city built around a particular factory or military base, where most people satisfy their instrumental desire for money by working on the same project.

In practice, most cities exist somewhere on a continuum between maximally diverse communities where the only shared interests are instrumental desires shaped by human biology and geographic proximity, and maximally specialized communities where everyone works in unison towards a specific end. Even in an urban area as diverse as Mumbai or New York, there are some social, cultural,

and political commonalities that are widely shared. Even in a highly specialized community like Los Alamos during the Manhattan project, or the Twin Oaks commune in Virginia, there is some disagreement about the best means to shared ends, as well as some individualistic desires to be with particular friends and family.

One thing to note however is that larger and more diverse cities tend to be more robust over time than smaller and more specialized ones. Cities built around common cultural values are at risk of cultural change causing most potential residents to choose to live elsewhere. Many former religious centers are now abandoned, once the religion was replaced. Cities built around a single factory or even a broader industry are at risk of external competition, technological change, or a change in resource availability causing this industry to dry up. Many former mining communities are now ghost towns, and manufacturing centers around the world have gone through boom and bust cycles as foreign markets first emerge and then later replace the center.

Meanwhile, cities that have more diversity, while they have a thinner shared agency, are more robust to change. A city that is not based on one industry is more likely to survive changes in importance of that industry. A city that is not organized around the principles of one culture is more likely to survive replacement of the culture. Major trading hubs of a multinational empire have often survived the downfall of the empire itself, because their multicultural interactions were able to continue through the new regime. This applies to the Aztec capital of Tenochtitlan that became Mexico City, and the Byzantine capital of Constantinople that became Turkish Istanbul, and even the Greek port of Massalia that became Roman and then French Marseille.

This suggests that even when a city has greater shared interests, due to specialization or similarity, there may be reasons for the city to focus its efforts on enabling the sorts of desires that are shared broadly based on common humanity. Even if there is currently shared culture or a central industry in the city, we know that these can often change, even within a single lifetime. If a city can devote its collective activities towards building infrastructure that helps with the more contingently shared desires (giving tax breaks for the film industry, or relaxing regulations on automotive manufacturers), or instead that helps more universally shared ones (standardizing labor safety conditions for all employers), it will likely better serve the long-run interests of its residents if it focuses on the latter.

There may be some policy implications of this suggestion. The transit consultant Jarrett Walker argues in his book *Human Transit* that the value of transit like buses and trains over cars is that they can carry more people through the same amount of space — for any dense concentration of people, this means a more efficient use of the limited resource of space. Furthermore, he argues that the traditional focus of urban transit systems on radial access to a single downtown core, with morning inbound focus and afternoon outbound focus, contributes most to serving just one type of desire (getting to work, for a certain type of employee). Instead, he argues that transit systems should be based around all-day travel over the whole high density area, to enable trips to visit

friends, go to school, go shopping, and do all the other many different things that people do other than commuting to work. While some cities may only have density in a single core, many modern cities have a large enough area with sufficiently high density for transit that it is more efficient to serve these many different types of trips through a frequent grid where most riders make one connection. (Walker, 2011)

Where some transit planners argue that cities should focus on the features of a transit system that make it pleasant, Walker argues that cities should focus on the features of a transit system that make it useful for many different types of people. Preference for rail transit over bus transit may have cultural components to it, but the advantages of cost-effectiveness and greater coverage of time and space are more enduring. (Walker, 2018) We might note that this sort of focus on the utilitarian virtues of public systems often seems aesthetically unpleasant. People often say that a particular road is unpleasant to drive on, or that a bus is unpleasant to ride, or that a restaurant is disgusting. But if the road or bus or restaurant is crowded anyway, this is a sign that it must be doing a very good job at being the most convenient or direct route to a commonly desired destination, or the cheapest and quickest option for eating.

4.2 Individual differences

One question for this picture of collective rationality for the city is how to deal with the fact that most of these interests I have mentioned are broadly shared but not totally universal. In some cases this is straightforward — even if a resident of Los Angeles never wants to venture north of the Hollywood Hills, she still has an interest in there being good transportation links, because so many of the businesses, government entities, and cultural institutions she cares about depend on commuters and visitors from the San Fernando Valley. There is a general sort of functioning of an urban area that is helpful to most of the interests that most residents could have, and this is what I think a city generally ought to promote, if it is to be rational.

However, there are likely some residents who have extremely divergent interests and needs, and it's less clear what my account of a city's rationality says about them. Are they not properly considered part of the city? Are their interests not relevant to considering rationality? Or are they relevant only to the degree that they do still share some interests?

One relevant case is that of extreme self-segregation. Ordinary segregation of racial and income groups produces some different priorities in terms of which transport connections are important, and which beautification and pollution mitigation programs are most important. But there is still a lot of shared interest. But some religious communities segregate to an even greater degree. The Satmar Hasidic community of Brooklyn is much more self-contained than most ethnic communities in the world. The Branch Davidian community that existed briefly in Waco was even more self-contained. Perhaps it is plausible to say that these communities are not really part of the city, because their specialized interests cut them off from the types of shared interests that people

living in geographic proximity with one another usually share. But they still share an interest at least in clean air and groundwater, even if they don't share municipal transportation and service infrastructure.

The interests most broadly shared are those due to agency, geographic proximity, and human embodiment. Thus, it is plausible that people with significantly different embodiment may have different interests. If certain abilities or disabilities give rise to different enough interests, then a person may no longer share enough with others to count as part of the collective whose shared goals I say are constitutive of rationality for the city. While people who never physically leave home may still desire the same transportation infrastructure to enable frequent visits from friends, family, and caretakers, there might be greater divergences of interest among people that use particular sorts of mobility aids that benefit from personal vehicle access to all spaces and would thus not benefit from pedestrian zones and public transit.

However, some interests that are stronger for people with particular disabilities are in fact broadly shared. The concept of "universal design" (Goldsmith, 1963, 2000) has led to widespread use of curb ramps at pedestrian crossings of streets, which turns out to greatly help children, people with bikes, and people carrying large objects, as well as people in wheelchairs and people with impaired joint mobility. Having auditory as well as visual signals at traffic intersections helps blind people cross the street, but also helps anyone whose visual attention might be distracted before the signal turns. In these cases, the shared desires were not obvious enough to have been implemented until the city started properly taking into account some of the apparently specialized interests of its residents with disabilities. Broadening the class of residents whose interests are considered has led to consideration of a broader class of interests shared by all, leading to collective behaviors that are more rational even when evaluated only with respect to the interests of the traditionally-abled.

5   (Dis)analogies with the Individual Case

One might worry that the sort of agency I ascribe to the city is too different from the sort that human individuals exhibit for the same characterization of rationality to be useful. This appears in certain distinctive ways that cities can fail to exhibit my notion of rationality. However, I claim that many of these examples have natural parallels in the individual case. Thus, if these are challenges to my application of this notion of rationality, then they are challenges in the individual case as well. While I don't have full responses to these challenges, I think that considering them can help illuminate some issues of rationality and agency both at the city level and the individual level.

Notions of rationality are often considered to be either explanatory (showing why certain behavior occurred) or predictive (giving a guide to what sort of behavior is to be expected). My account is not quite either — it is instead intended to evaluate how well some individual or community is doing at living up to its own goals. I don't need to posit that there is some character trait

of rationality that persists through time, to explain or predict behavior of the more or less rational agents. I just want to be able to characterize particular behaviors or habits as rational or not, and not the agent as a whole (whether a city or a human).

One might think that there are at least two different types of irrationality that a city can exhibit on my account. It might exhibit "normal" irrationality, where it just does something against its interest, like voting to demolish a downtown block for a football stadium that will only be used on a dozen nights of the year. Or it might exhibit "divided" irrationality where one faction within the city wins out against others, as in exclusionary zoning that helps the middle class and hurts the working class. However, I think this "divided" irrationality is a lot like akrasia for the individual. Both my present self and my tomorrow self care about various long-range plans, that might all best be served by being alert and awake tomorrow. However, my present self also has interests in having fun at the party tonight, and my tomorrow self has extra interest in avoiding the unpleasantness of a hangover. It may be that my present self wins out, even harming some of its own interests in order to satisfy its strongest interest.

This emphasizes the worry raised earlier that agency of the sort I ascribe to the city exists also at higher and lower levels, like the neighborhood or nation, as well as ethnic and socioeconomic groupings both within and across cities. I claimed that the city is a particularly distinctive level where a large number of shared interests coincide. But the individual level is also only one particularly natural level of analysis, among many nearby ones. We contain multitudes, and while their interests often align, they don't always align exactly. As a sibling, a professor, a romantic partner, and a citizen I have many different interests, some competing and some aligning. My present self and my future selves have many interests in common and many divergent interests. Simon Evnine (2003, 2005) claims that the time-slice is a distinctive level, but it seems that even there, there are cross-cutting interests and disunified beliefs. (Lewis, 1982)

In my account of rationality for the city, I discussed legal and physical infrastructure, as well as collective behaviors. It might seem that behaviors are the only analog at the level of the individual. Furthermore, some infrastructure that seems irrational now, like New York's combined sewer system, or Boston's irregular street pattern, may have been rational at the time it was built, and now the costs of changing it would be too great to be worthwhile. However, I claim that this is only because philosophers have often neglected the sort of infrastructure that is essential for the individual as well. In our daily lives, we are aware that many of our behaviors are directed towards providing the infrastructure to enable more successful future behavior. I try to exercise regularly and avoid sitting for long periods of time partly because I enjoy the feeling of being healthy, but also partly because I hope that this will provide the physical fitness that will help me achieve my goals for decades to come. My students attend my classes hopefully partly for their present edification, but also in part to provide the mental and conceptual infrastructure (and social certification) that will help them achieve their goals in the future. And just as Boston and New York have limitations on their current capacities for dealing with transportation and water

due to past decisions that stuck them with poor infrastructure, there are many people with present limitations on their capacities for achieving their personal goals as a result of past decisions that resulted in broken bones or unfortunate tattoos, or due to scars that resulted from surgeries performed at a time when more invasive techniques were used. Education and physical training are the personal analogs of investment in civic organization and infrastructure.

6   Conclusion

I have claimed that for any notion of shared desire that can shape shared behavior, there is a corresponding notion of shared agency and shared rationality. I have argued that this notion of agency and rationality is particularly pronounced at the level of the city (considered as an entire urban area, more than a municipality), and that we can therefore apply intuitive notions of rationality to social norms and the collective behavior that gave rise to the legal and physical infrastructure of various cities. Furthermore, I have argued that this notion of rationality recommends that cities act in ways that promotes the instrumental interests that all humans tend to share in virtue of their shared embodiment and geographic proximity, more than the interests that residents of a given city happen to share beyond that.

A natural question arises about whether this notion of rationality for a city corresponds to any notion of justice or morality. I have shied away from this question. In many instances rationality does seem to correspond to some notion of justice. A city that is rational in my sense will provide clean air and water for its residents, and public transportation infrastructure to help them get around regardless of their individual wealth or resources. It will try to help its residents get out of poverty, because the presence of poverty tends to harm the interests of neighbors as well as those that are poor.

But there are also cases in which my notion of rationality doesn't seem to correspond to justice. If some disabilities require interventions that go beyond universal design, we may get conflicts there. And interventions that have made cities more rational in my sense have often done poorly by many measures of justice. Paris is the beloved and successful city it is today in part because of the broad boulevards lined with Beaux Arts-style buildings that connect the neighborhoods together. But to put these in, Baron Hausmann under Emperor Napoleon III demolished many neighborhoods of poor people. New York's Central Park is a major feature that benefits all current residents. But it required the expropriation of black neighborhoods of freed slaves. Reformers that want to benefit the schools or transit system or housing supply of modern cities often have to confront the fact that their plans would require the devastation of many individual residents. (And conversely, preservationists have to confront the fact that their plans require the continual failure and exclusion of the systems for those that don't already benefit from the current systems.) I don't want to commit myself to the claim that these interventions are just, even if they do fit my notion of rationality. On the other hand, it may turn out that even by my

notion of rationality these sorts of projects don't meet the mark — the urban renewal projects of the 1960's in the United States that were meant to ease transportation and benefit residents by removing blight ended up in fact causing massive traffic congestion and hollowing out of the tax base of the cities, in addition to the acknowledged costs of dislocation during their implementation.

However, again I think this challenge reveals an interesting parallel at the level of individual humans that isn't often considered. If sub-personal levels of agency have some sort of moral standing, then many personally rational decisions may be morally problematic as well. We tend to think of self-discipline and self-control as unalloyed goods, but this may be because we are ignoring the interests of our sub-personal selves. Spending one's youth training for adulthood can be successful at its goals, but it can come at the cost of one's youthful enjoyment. If the end result of the educational process is a person whose values differ from the person that went in, then it's hard to say that the person who pays the costs is the person who gets the benefits later on — there's also a sense in which the person who paid the costs is just extinguished and replaced with someone else, in a kind of gentrification of the soul. There are surely problems with this analogy in both directions, but I hope that it can lead to productive new ways of conceptualizing these ways of developing over time.

References

Alexander, C. (1965). A city is not a tree. *Architectural Forum*, 122(1):58–62.

Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy*, 100(7):356–371.

Aron, H. (2017). A guide to L.A. etiquette in 6 public spaces. *LA Weekly*.

Bernstein, S. (2003). Turning left on red is part of driving culture. *The Los Angeles Times*.

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*.

Bradley, D. and Leitgeb, H. (2006). When betting odds and credences come apart: more worries for Dutch book arguments. *Analysis*, 66(290):119–127.

Christensen, D. (1991). Clever bookies and coherent beliefs. *The Philosophical Review*, 100(2):229–247.

Cohen, A. (2008). The ultimate Kantian experience: Kant on dinner parties. *History of Philosophy Quarterly*, 25(4):315–336.

Easwaran, K. (2011). Bayesianism I: Introduction and arguments in favor. *Philosophy Compass*, 6(5):312–320.

Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(2):143–147.

Evnine, S. (2003). Epistemic unities. *Erkenntnis*, 59:365–388.

Evnine, S. (2005). Containing multitudes: Reflection, expertise, and persons as groups. *Episteme*, pages 57–64.

Gilbert, M. (1990). Walking together: a paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15(1):1–14.

Gillies, D. (1991). Intersubjective probability and confirmation theory. *British Journal for the Philosophy of Science*, 42:513–533.

Goldsmith, S. (1963). *Designing for the Disabled: A Manual of Technical Information.* Royal Institute of British Architects.

Goldsmith, S. (2000). *Universal Design: A Manual of Practical Guidance for Architects.* Architectural Press.

Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459):607–632.

Hedden, B. (2015). *Reasons without Persons.* Oxford University Press.

Hitchcock, C. (2004). Beauty and the bets. *Synthese*, 139(3):405–420.

Horowitz, S. and Dogramaci, S. (2016). Uniqueness: a new argument. *Philosophical Issues*, 26.

Kolodny, N. (2007). How does coherence matter? *Proceedings of the Aristotelian Society*, 107:229–263.

Kopec, M. (2015). A new group dutch book argument. *Ratio*, 30(2):122–136.

Kopec, M. and Titelbaum, M. (2016). The uniqueness thesis. *Philosophy Compass*.

Lewis, D. (1982). Logic for equivocators. *Noûs*, 16(3):431–441.

Moss, S. (2015). Time-slice epistemology and action under indeterminacy. *Oxford Studies in Epistemology*.

Muldoon, R., Lisciandra, C., Bicchieri, C., Hartmann, S., and Sprenger, J. (2013). On the emergence of descriptive norms. *Politics, Philosophy & Economics*, 13(1):3–22.

Omohundro, S. (2008a). The basic AI drives. In Wang, P., Goertzel, B., and Franklin, S., editors, *Proceedings of the First AGI Conference*, volume 171 of *Frontiers in Artificial Intelligence and Applications*.

Omohundro, S. (2008b). The nature of self-improving artificial intelligence. unpublished.

Paul, L. (2014). *Transformative Experience.* Oxford University Press.

Pierce, J. and Kolden, C. (2015). The hilliness of U.S. cities. *Geographical Review*, 105(4):581–600.

Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48(2):193–218.

Searle, J. (1990). Collective intentions and actions. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communication*, pages 401–415. MIT Press.

Talbott, W. J. (1991). Two principles of Bayesian epistemology. *Philosophical Studies*, 62(2):135–150.

Teller, P. (1973). Conditionalization and observation. *Synthese*, 26:218–258.

Tollefsen, D. (2002). Collective intentionality. *Internet Encyclopedia of Philosophy*.

Walker, J. (2011). *Human Transit*. Island Press.

Walker, J. (2018). To predict with confidence, plan for freedom. *Journal of Public Transportation*, 21(1):119–127.

Whyte, W. H. (1980). *The Social Life of Small Urban Spaces*. Project for Public Spaces.