SUPPLEMENTAL MATERIAL

Methods

Data Analysis

Data were analyzed using R (Version 3.3.1; R Core Team, 2016), R Studio (Version 1.0.136), Python (van Rossum and Drake, 2001), and depthmapX (Version 0.5b; Varoudis, 2015b). Points (x,y) defining the maze boundaries were imported into Adobe Illustrator's graph tool, converted to lines, and then exported to .dxf format. The maze and human data were imported into depthmapX (Figure 1B), and then binned at a variety of spatial resolutions.

Spatial binning. Because the walls of the maze corridors in the present study were 1.25m wide, employing bin sizes larger than 1.0m did not produce sensible results (e.g., a bin size of 1.1m produced cells that straddled both maze corridors and the inaccessible spaces between corridors). Due to the combinatorial explosion associated with computing space syntax measures at very high spatial resolution, it was impractical to compute space syntax measures for bin sizes smaller than .01 meters (3.94 inches).

Agent analysis. Default depthmapX settings for agent analysis were used, with the exception of a few parameters that were set to approximate those of the experimental design (for a detailed discussion of setting parameters for agent-based analysis using visibility graphs, see Turner 2003). First, the "Analysis length (timesteps)" parameter was set to 43,200 in order to approximate the parameters of the exploration phase (12 mins of exploration x 60 seconds per min x 60 Hz walking data sampling rate). Agents were released from a position at the center of the maze roughly corresponding to the location at which participants began the experiment. Finally, "Record trails for N" agents was set to N = 36, corresponding to the number of participants in the experiment. As Turner (2003) notes, these agent based analyses "approximate a Markov chain operating through locations on the visibility graph."

Data alignment and correlation method. Due to minor inconsistencies between depthmapX's coordinate system and the coordinate system employed in R (which ranged from -10 to 10), a custom R script was used to align the configurational bins and binned pedestrian count data. Matrices containing binned walking data (Figure 1C) and syntactic measures computed with the same spatial resolution (Figure 1D) were superimposed and then systematically shifted (up, down, left, and right) until an optimal overlap was found using a least-squares criterion (maximum Pearson's product moment R^2 value).

Data Transformation Analysis: Examining whether (Q1) correlations are sensitive to data transformations. Silva (2013) explicitly recommends log transforming pedestrian movement data to ensure that both movement and syntactic data follow normal distributions, enabling statistical comparisons between them. Exploratory data analysis suggested a variety of possible related data transformations beyond those recommended by Silva (2013), so we decided to systematically examine the impact of additional data transformations on correlation strength. First, raw walking data (W) values were correlated with raw syntactic (S) values (W vs S). Second, because it was possible to obtain syntactic or walking data values of zero, the analysis was restricted to values that were greater than 0 for both the walking data and space syntax data (*W>0 vs S>0*). Third, the natural logarithm of values produced using the previous method was also examined. Finally, because $\log_{10}(0)$ and ln(0) are undefined, and because the log+1 transformation is commonly used to correct for departures from normality, log+1 transformed walking data was also compared to raw syntactic values [e.g., *log(W+1) vs. S*, and *ln(W+1) vs. S*].

Regression Analysis: Examining whether (Q2) correlations depend on the spatial resolution of the sampling grid. We wished to be conservative in testing (Q2) whether correlations would decrease with increased spatial resolution. This required identifying data transformations that would be most charitable (i.e., that would allow space syntax measures the greatest chance to remain high as we increased the spatial resolution of the underlying sampling grid) toward a wide variety of measures and spatial resolutions (bin sizes). To accomplish this, two complementary approaches were taken. First (Part 1), we identified and opportunistically applied whichever data transformation (of the 11 transformations examined) produced the highest correlation for a given measure-bin pair. Second (Part 2), we identified a single (best overall) data transformation that produced maximal correlations for the largest percentage of measure-bin size pairs (see Table 2).

Simulations: Examining whether (Q3) a small sample of spatial locations would yield spuriously high correlations. Past research has generally sampled pedestrian data at subsets of locations ("gates" or grid cells) within the overall VGA sampling grid, rather than sampling pedestrian data at all possible sampling grid locations. Historically, the number of "gate" locations (N_{GATES}) has been limited due to data collection constraints (e.g., needing large numbers of researchers to collect data, or relative ease of counting pedestrians passing through doorways), and because gates are often positioned at locations convenient for researchers, it is possible that high correlations obtained in previous studies may be due to selection bias. In contrast, the motion tracking system used in the present study recorded *all* possible locations within the sampling grid, providing a more comprehensive assessment of syntactic predictions. In addition, we examined how correlations vary as increasingly large *subsets* of grid cells are randomly selected, simulating how stationing an increasing number of randomly located “experimenters” (N_{GATES}) to record gate counts impacts correlations. We evaluated whether using small subsets (n = 3) of grid cells to compute R^2 values would yield spuriously high correlations, and whether sampling from an increasing number of locations (up to n = 100) would yield more reliable or “stable” correlations (denoted R^2_S).

Several approaches were used to quantify how syntactic-behavioral correlations (R^2 values) vary as the number of “gates” (N_{GATES}) is increased. First, 100 replications (R^2 values) were computed for each simulated value of N_{GATES} (this corresponds to randomly distributing 100 distinct sets of N “experimenters” to count pedestrian flows for each N_{GATES} value, where N = N_{GATES}), yielding a total of 100,000 simulated R^2 values [(100 gates) * (100 replications/gate) * (10 bin sizes)] for each syntactic measure. Exploratory data analysis suggested that (a) R^2 values were highest when the number of sampling grid locations (N_{GATES}) was relatively low, and that (b) the mean R^2 value appeared to decline exponentially, before stabilizing above a critical value of N_{GATES}. Therefore, a change point approach was used quantify the presence of inflections or “change points” (CP, the gate count at which correlations tended to stabilize) in simulation data; thus, CP is the critical measure used to assess the minimum number of sampling grid locations (N_{GATES}) required to obtain reliable estimates of correlation strength at a given grid resolution.

Change points were detected by first computing local polynomial regression (LPR) fits for simulated R^2 values as a function of N_{GATES} (using the “loess” function from R’s “stats” package), and then obtaining the first detected change point in these regression fits (using R’s “cpm” package) (Ross, 2015). The initial LPR fitted the R^2 value at the minimum number of gates examined (N_{GATES} = 3), indicated by R^2_I. The arithmetic mean of the LPR fitted R^2 values between N_{GATES} = CP and N_{GATES} = 100 was used to estimate the point at which R^2 stabilized at a relatively constant value (R^2_S). Finally, the difference between the two R^2 values ($\Delta R^2 = R^2_I - R^2_S$) was computed to examine how correlations vary as the number of sampling grid locations (N_{GATES}) was increased.

Results

Q2: Do correlations depend on the spatial resolution of the sampling grid?

Heatmaps for Leading Syntactic Measures

This section examines whether (Q2) correlations depend on the spatial resolution of the sampling grid by discussing data for a syntactic measure that generally performs well in the space syntax literature [Visual Integration (R3)], and for the syntactic measure that yielded the highest correlation obtained in the present study [Metric Node Count (R1)]. Heatmaps in Figures 5 and 6 plot syntactic values and binned walking data at corresponding grid resolutions for Visual Integration (R3) and Metric Node Count (R1) respectively. Heatmap values and correlations were computed at each bin size after applying a $Log_{10}(W)>0$ *vs. S* data transformation; color scales indicate data ranges and color mappings.

Heatmaps and Correlations for Visual Integration (R3). As previously noted, Integration is a commonly reported measure in the space syntax literature. Visualizations of Visual Integration (R3) and binned walking data for all 10 bin sizes appear in Figure 5.

[Insert Figure 5]

For this measure, R^2 values increased from R^2 = .17 at the lowest spatial resolution (1.0m) to a maximum of R^2 = .37 at an intermediate resolution (0.7m), and decreased to R^2 = .13 at the highest spatial resolution (0.1m) examined. Thus, we found that (Q2) correlation strength decreased (by .05) as spatial resolution was increased. It is worth nothing that R^2 values reached a peak at $R^2 = .37$, which may provide support for the notion that there is an ideal resolution for syntactic measures (Al Sayed et al., 2014; Turner et al., 2001).

Heatmaps and Correlations for Metric Node Count (R1). Visualizations of Metric Node Count (R1) and binned walking data for all 10 bin sizes appear in Figure 6.

[Insert Figure 6]

Metric Node Count (R1) yielded the highest correlation value found in the present study. For this measure, R^2 values increased from $R^2 = .18$ at the lowest spatial resolution (1.0m) to a maximum of $R^2 = .41$ at an intermediate resolution (0.6m), and decreased to $R^2 = .19$ at the highest spatial resolution (0.1m) examined. With respect to Q2, this syntactic measure exhibited a more complex pattern of results than we predicted, with correlations peaking at intermediate bin sizes (see Figures 2 and 3), which may be consistent with the claim that there is an ideal spatial scale for computing syntactic-behavioral correlations (Al Sayed et al., 2014; Turner et al., 2001).

Q3: Does a small sample of spatial locations yield spuriously high correlations?

Each boxplot (Figures 4 and 7) shows simulated R^2 values (y-axis) against the number of randomly sampled gate locations (x-axis N_{GATES}) for a given bin size. Boxes and whiskers summarize the distribution of the results from all 100 replications at each value of N_{GATES}; whiskers extend to the minimum and maximum simulated R^2 values, and extend no further than 1.5 times the interquartile range (IQR); box hinges indicate the 25^{th} and 75^{th} percentiles of the simulated R^2 values; outlying points are indicated as black dots.

Metric Node Count (R1). At all ten spatial resolutions examined, correlations between Metric Node Count (R1) and walking data decreased (ΔR^2; $M = -.18$, $SD = .042$) as N_{GATES} increased. The first value of N_{GATES} (Figure 7, x-axis) at which a significant change (Ross, 2015) in LPR fitted R^2 values (Figure 7, y-axis, blue best fit line) was detected was $N_{GATES} = 23$, just as we found for Visual Integration (R3).

[Insert Figure 7]

This value was consistent across all ten of the spatial resolutions examined. Beyond 23 gates, correlations tended to stabilize (R^2_S) at a low but relatively constant value (*mean* $R^2_S = .258$; $SD = .151$). With respect to Q3, when fewer than 23 gates were used to compute correlations, perfect positive correlations ($R^2 = 1$) between random noise and walking data were obtained, strongly suggesting that using a small number of sampling grid locations can inflate correlations.

Comparisons to random noise. To assess whether this measure correlated with walking data above chance levels, random noise was substituted for syntactic data, and correlated with walking data. Initial correlations (R^2_I) between syntactic data and walking data (*mean* $R^2_I = .43$, $SD = .13$) were 72% higher than correlations between random noise and walking data (*mean* $R^2_I = .25$, $SD = .01$), $t(9) = 4.37$, $p < .01$. Stabilized correlations (R^2_S; beyond $N_{GATES} = 23$) with walking data were also higher for syntactic data ($M = .257$, $SD = .15$) than random noise data ($M = .02$, $SD = .001$), $t(9) = 4.98$, $p < .001$. Thus, syntactic measures performed better than chance. However, with respect to Q3, when fewer than 23 gates were used to compute correlations, perfect positive correlations ($R^2 = 1$) between random noise and walking data were obtained, strongly suggesting that using a small number of sampling grid locations can inflate correlations.

Discussion

Local maxima. Some measures exhibited a small "hump" or local maximum in correlation strength (R^2) at an intermediate spatial resolution near 0.7m (see Figures 2 and 3). In an effort to adopt a scale of analysis commensurate with typical human walking behavior, Turner et al. (2001) employed a 1m grid spacing, and Al Sayed et al. (2014; *depthmapX* handbook) recommends that researchers depthmapX users select “a sensible grid spacing values that match the human scale (0.6 - 0.7 meters).” Thus, our results could be interpreted as supporting the claim that there is an optimal spatial scale for correlating space syntax measures with pedestrian behavior; the large number of syntactic measures examined in the present seems to support the recommendations made by other researchers. However, we urge caution with respect to this interpretation of our results. While several previous studies (Emo et al., 2012; Ferguson et al., 2012; Turner, 2003) cite Gibson’s (1950, 1986) ecological approach to visual perception as the theoretical basis for positing a causal relationship between syntactic variables and pedestrian behavior, they do not clearly articulate why syntactic-behavioral correlations should be maximal at human scale. Moreover, operational definitions of “human scale” have been extensively debated, and remain controversial (see Ewing & Handy, 2009 for a review).

Table 2
Summary of syntactic-behavioral correlations found in selected previous studies

Study	Mode	Environment	Syntactic Measure(s)	Data Transformations	Correlations
Hillier et al. (1996)	Walking	Museum	Integration	ln (movement rates)	$.37 < R^2 < .86$
de Arruda Campos (1997)	Walking	Urban area	Integration (R3) Integration (RN)	(Unknown)	$.81 < R^2 < .88$ $.80 < R^2 < .80$
Penn, Hillier, Banister, and Xu (1998)	Vehicle	Urban area	Integration (R3, R5, R7, R9)	Fourth root of flow rates	$.34 < R^2 < .83$
	Walking	Urban area	Mean integration (R3) and development density	Net capacity	$R^2 = .98$
Turner & Penn (1999)	Walking	Museum	Isovist Integration	Log of mean occupancy levels	$R^2 = .585$
		Store	Isovist Area		$.324 < R^2 = .578$
Desyllas & Duxbury (2001)	Walking (5m and 3m)	Urban area	Axial Map Analysis	ln (mean visibility) and ln (mean pedestrian movement data)	$R^2 = .456$ (5m) $R^2 = .625$ (3m)
Turner (2003)	Walking (3m)	Urban area	Various	Log transformed agent simulation data	$.29 < R^2 < .73$
Turner (2007)	Walking	Museum	Through vision (agent simulation)	ln (movement rates)	$.68 < R^2 < .74$
Mora, Astudillo, and Bravo (2014)	Walking	Urban area	Gate counts ($N_{GATES} = 203$)	Mean gate counts over six consecutive workdays	$.142 < R^2 < .271$
Okamoto et al. (2013)	Walking	Commuter rail mall	Gate counts ($N_{GATES} = 50$)	Connectivity, visual step depth, shortest distance, integration	$.2 < R^2 < .598$

Note: Mode column indicates whether pedestrian (walking) data or vehicular data were correlated with syntactic measure(s). Correlations column includes minimum and maximum syntactic-behavioral correlations found in the study. N_{GATES} = the reported number of gate locations at which pedestrian flows were counted.

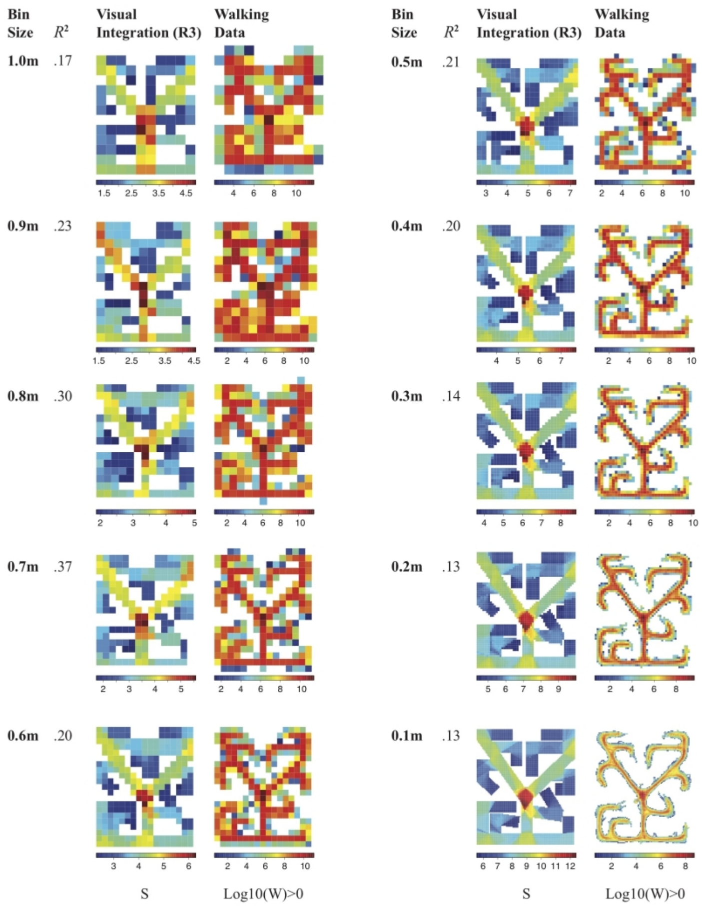

Figure 5. Heatmaps and Correlations for Visual Integration (R3). Heatmap values and correlations were computed at each bin size after applying a Log10(W)>0 vs. S data transformation. For Visual Integration (R3), colors indicate raw syntactic measure values (S). As previously noted, the Integration value for a grid cell is obtained by computing the average depth (i.e., topological distance) of that cell to neighboring cells within a specified topological distance (radius), effectively ranking cells "from the most integrated to the most segregated" (Klarqvist, 1993). For the walking data, colors indicate Log10(W)>0 transformed position data count values computed for each grid cell; the same walking data are plotted in both Figures 5 and 6.

215x279mm (300 x 300 DPI)

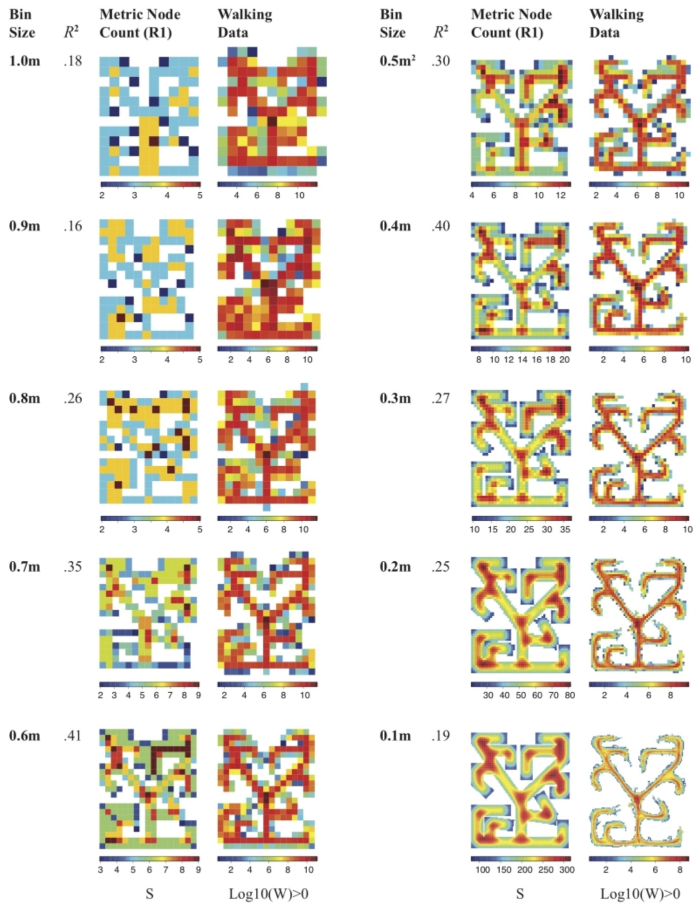

Figure 6. Heatmaps and Correlations for Metric Node Count (R1). Heatmap values and correlations were computed at each bin size after applying a Log10(W)>0 vs. S data transformation. For Metric Node Count (R1), colors indicate raw syntactic measure values (S). Metric Node Count (R1) is the number of neighboring nodes within a specified topological distance (radius) (Turner, 2004). For the walking data, colors indicate Log10(W)>0 transformed position data count values computed for each grid cell; the same walking data are plotted in both Figures 5 and 6.

215x279mm (300 x 300 DPI)

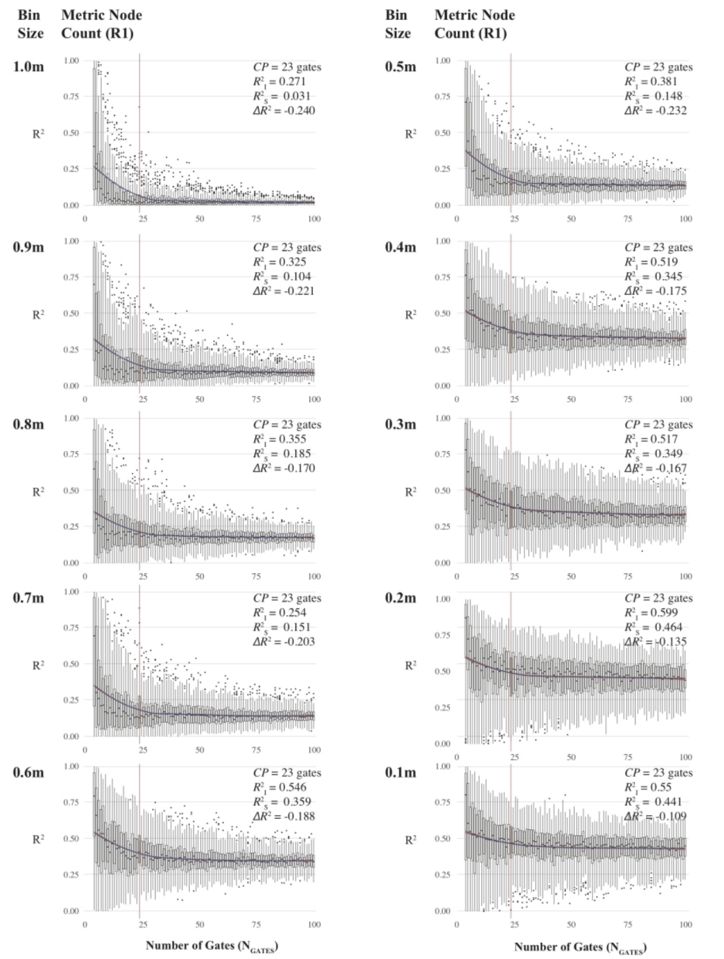

Figure 7. Simulations: Metric Node Count (R1). *x-axes* (N_{GATES}): number of randomly sampled gate locations; 100 replications were simulated for each value of N_{GATES}. *y-axes*: R^2 values between the 100 simulations and the walking data. The mean and range are displayed for each value of N_{GATES}. *Boxplot whiskers*: min/max of 1.5x interquartile range. *Black dots*: outliers. *Trend lines*: best fit line for local polynomial regression (LPR) fit. *CP (red vertical line)*: first estimated change point value in LPR fitted R^2 values. R^2_I: initial R^2 value of LPR fit line (at $N_{GATES} = 3$). R^2_S: stabilized R^2 value, estimated by obtaining the mean of all LPR fitted R^2 from N_{GATES} = CP gates to $N_{GATES} = 100$. $\Delta R^2 = R^2_I - R^2_S$. For each value of N_{GATES}, 100 simulation runs were performed. *Boxplot hinges*: 25th and 75th percentiles.

215x279mm (300 x 300 DPI)