APPENDIX

The approach taken here simulated the reverse of what Townsley and Sidebottom (2010) demonstrated. Starting with simulated distance sets where each set had 15 distances and each set of values was generated to represent a normal Gaussian distribution, adding just *one* extreme outlier, which at the *most* was 100 percent greater than the mean distance before it was substituted in, and at the *most* 10 standard deviations from the mean before it was substituted in, resulted in a high proportion of distance sets with significant skewness and non-normality.

Townsley and Sidebottom (2010) observed for both types of burglary targets that, among burglars with 10 or more detections, the set used for their Figure 3, the most typical burglar was in the category of 10-20 detections over the ten year period. Taking the middle – 15 – of that most frequent interval, 100 sets of normally distributed distances for 100 hypothetical burglars were generated. For each of the 100 sets of 15 distances each, a mean was randomly selected between 1 and 2 in increments of .10, and a standard deviation was randomly selected ranging from .10 to .25 in increments of .05. It was necessary to make the standard deviation small relative to the mean values so as to avoid negative values, an impossibility if distance is the variable in question. Using a normal sample generator (Stata's drawnorm), and a different randomly chosen 5 digit random seed for each set, 100 sets of distances were generated representing the hypothetical distance sets for 100 burglars. No minima fell below .5, no maxima were above 2.6.

As would be expected just due to chance, five percent of the distributions had significant skewness ($p < .05$), and six percent of the distributions had significant ($p < .05$) non-normality as indicated in the D'Agostino-Pearson K^2 omnibus test for normality (D'Agostino, et al., 1990). This omnibus test considers both moments $\sqrt{\beta_1}$ and β_2 reflecting, respectively, skewness and kurtosis.

One value was randomly selected and removed from each series, and replaced with an outlying value of 2.0. Doing so resulted in 37 percent of the distance series having significant skewness ($p < .05$). The same percent of distance sets were significantly ($p < .05$) non-normal as indicated by the omnibus test.

Following Townsley and Sidebottom's (2010) approach, if one did not know that the initial distance sets were normally distributed, one would conclude at this point that 37 percent of the burglars' trips reflected the distance-decay function, and 63 percent did not.

If, instead of inserting a value of 2.0, a value of 2.5 was used, then 87 percent of the distance series had significant skewness ($p < .05$), and a significant K^2 statistic indicating non-normality. If one did not know that the original distributions were generated to match a normal distribution, one might conclude that 87 percent of the burglars' trips reflected a distance decay function, and 13 percent did not.

Clearly results such as this depend on the relationship between the standard deviation and the mean, and the number of observations in the distance set. Changing any of these parameters would obviously alter the results observed here. The only point being suggested here is that although using the significance of skewness represents one way to discriminate between normal vs. non-normal distributions, with the latter presumably following a distance-decay function, it has limitations. Because the skewness ($\sqrt{\beta_1}$) numerator relies on *cubed* values of the deviation between a score and the mean, one extreme outlier makes a normal distribution non-normal when the n of cases in the set is relatively small. A normal distribution with one extreme outlier is not equivalent to a distribution following a distance-decay function where the frequency decreases with the log of the distance.

A more specified approach (de Vries, Nijkamp, & Rietveld, 2009) may be warranted for identifying whether a distance decay function is present, and if so, which one.