**Appendix I. Detailed description of the model**

To explore the impact of preferential intersection on paleobiological sampling, we developed a simple model that captures the basic aspects of fossil sampling from lithified rock. The model in its simplest form consisted of a cubic volume with some number of fossil specimens distributed in the volume. This populated volume, which was analogous to a fossiliferous rock, was then bisected by one or more horizontal planes, corresponding to planar fractures, at random points along its vertical axis. Specimens that intersected one or more planes were considered exposed and were counted, contributing to the fossil sample. This process was then repeated until a user-defined sample size quota was reached.

In order to represent in Matlab the components of this model (i.e., the cubic volume, the bisecting planes, and the specimens distributed in the volume), they were reduced to simplified geometric forms that could be represented as sets of coordinates. This was straightforward for the rock volume and fracture surfaces. The rock volume was represented by eight sets of coordinates that together define the corners of a cubic volume. A single point on the vertical axis defined the two-dimensional plane that bisected this volume. The fossil contents, which in our model were mollusk shells, were more geometrically complex than simple cubes or planes but could also be represented as simplified geometric forms. To do this, we used Raup (1966)’s coiling model (Fig. 1). Raup’s model is capable of replicating the shapes of a wide variety of molluscan shells using a parameterized equation. Briefly, Raup’s model consists of a set of starting points placed around a circle with lines starting from these points coil around a vertical axis. The morphology of the resulting shell is controlled by the rate of whorl expansion (i.e., the rate at which radius of the circle around which the points are placed increases in size with coiling) and translation rate (i.e., the rate by which the circle shifts along the vertical axis relative to coiling around the axis). Examples of model shells generated using our Matlab code written to implement Raup’s model are shown in Figure 1. Using Raup’s models we generated morphologies for a generic bivalve and gastropod that could then be modified to reflect differences between species in terms of size and shape and be distributed in the three-dimensional cubic volume. The Matlab implementation of our analog model therefore consisted of a cubic volume populated by specimens, all represented by sets of coordinates.

We first created nine parent populations of shells, from which specimens could be drawn at random to populate the cubic volumes. These nine parent populations represented different combinations of three species abundance distributions and three sources of species size data, which are described below. Each parent population contained 1,000,000 specimens to ensure an adequate sampling universe. These specimens represented different species that followed an abundance distribution. For simplicity, each species was defined by a distinct but uniform size and shape (i.e., there were inter-specific variations in size and shape, but there were no intra-specific variations). The model used shell height (i.e., distance from the dorsal to the ventral sides of specimens) as a measure of size.

Data on species abundances is commonly presented in the paleoecological literature, however these are usually limited to samples of only a few hundred or at best thousands of individuals. Therefore, in order to have a sampling universe sufficient to explore variation in diversity up to sample sizes of thousands of individuals, we chose to generate artificial datasets of size N = 1,000,000 based on realistic parameters. To do this, we used three species abundance distributions: two following the log series distribution (Fisher et al. 1943) and one following the lognormal distribution (Bulmer 1974). In one of the log series distributions, we set Fisher’s α equal to 6, which yielded species richness-sample size relationship that was similar to those observed in the studies of Sessa et al. (2009) and Hendy (2009). In the second log series distribution, we used a Fisher’s α value of 3, in order to examine the behavior of a less even and less species-rich assemblage. For a given α value and a set number of individuals (N = 1,000,000), the log series distribution predicts a specific species richness in the assemblage (i.e., 39 species when α = 3; 87 species when α = 6). These species richness values correspond to the number of species that would be recovered if all 1,000,000 individuals were sampled.

We followed the methods of Wagner et al. (2006) in generating lognormally distributed distributions. Unlike the log series distribution, the lognormal species abundance distribution does not specify a particular species richness associated with a particular set of parameters. Rather, a particular lognormal distribution specifies how species are distributed among abundance classes while the number of species within a particular lognormally distributed assemblage is a user-defined parameter. For our analyses, we generated lognormally distributed parent distributions with total species richness values of 50. These distributions yielded parent populations with richness values of the same order of magnitude as those produced by log series distributions. Parameter values for the lognormal distribution were μ = 1.2281 and σ = 1.2408.

Three sources of size data (Roy et al. 2000; Kowalewski and Hoffmeister 2003; Sessa et al. 2009) formed the basis for constructing the parent populations with varying size between model species. The supplementary dataset of Kowalewski and Hoffmeister (2003) provided lists of specimens identified to species level with associated measurements of shell height. The mean of individual species provided a basis for a dataset of species sizes. Roy et al. (2000) provided numbers of species assigned to different size bins. Sessa et al. (2009) provided lists of species taken from both her own work and previously published sources. While Sessa et al. (2009) did not provide species size information, one of her sources of species abundance data, Garvie (1996), gave species-level size information including shell heights, width:height ratios, and length:height ratios. Measurements were given for the largest specimen of a species. This information was used as a source of size information for the species lists of Sessa et al. (2009).

Data on shell height were available in Kowalewski and Hoffmeister (2003) and Sessa et al. (2009). However, Roy et al. (2000) recorded specimen size as the average of shell height and length. For the purposes of modeling, we chose to treat the Roy et al. (2000) size data as measurements of shell height. This means that the species size distribution referred to here as Roy et al. (2000) is not identical to what would have been produced had Roy et al. (2000) chosen to use only shell height as a measure of size. Since the Sessa et al. (2009) dataset covers a larger size range than the two other datasets, analysis of these three datasets is potentially informative about how the distribution of species among size bin impacts the results of preferential intersection, — a primary goal of this study.

We used the same binning scheme, log2, for all three size datasets. We then converted these binned species size datasets to lists of species size values for each of the three datasets. The number of times a size bin value occurred in a list corresponded to the number of species that fell in that size bin in that particular dataset. We used the midpoint between log2 values as the cutoff point to determine which size bin a species went in, i.e. a species with a height of 35 mm would be placed in the 32 mm size bin while a species with a height of 39 mm would be placed in the 45.3 mm size bin. The sizes assigned to species corresponded to the value of the size bin, i.e. all species placed in the 32 mm bin were assigned a size of 32 mm. These lists of binned sizes served as the basis for assigning size values to model species. The three binned size distributions are shown in Figure 4.

In contrast to specimen size, where we used three datasets to explore the effect of species size, only a single dataset on shell shape was used for all analyses. This dataset was generated by taking the species list of Sessa et al. (2009) and downloading data on measurements of shell height, width, and length from the Paleobiology database (retrieved May, 2016). The PBDB only presents measurements at the genus level, so the list of measurements was made at the genera level. Sets of measurements were then converted to aspect ratios by linearly scaling shell height to 1. For genera whose height-width-length data were not available in PBDB, their width:height and length:height ratios were taken from Garvie (1996).

The compiled data of relative abundance, size, and shape were then used to generate parent populations from which specimens could be drawn to populate model rock volumes. To do this, the code first created a 5 × Sp matrix, where Sp equals the number of species in a particular parent population (i.e., 39 for log series with α = 3; 87 for the log series with α = 6; and 50 for the lognormal dataset) and each column corresponds to a single species. The first row corresponds to species size and was filled by randomly selecting (without replacement) Sp values from the list of bin sizes. The second and third rows correspond to width:height and length:height ratios. These rows were similarly filled by randomly selecting (without replacement) Sp entries from the list of aspect ratios. The fourth row was set equal to species abundance values generated by the log series or lognormal species abundance code. Entries in the fifth row were randomly assigned either 1 or 0, corresponding to bivalve or gastropod identity, respectively. Two thirds of species were labelled as gastropods while the rest were labelled as bivalves, similar to the ratio observed in the data of Kowalewski and Hoffmeister (2003). The sixth row contained ordered number entries from 1 to Sp, recording species identify. This 5 × Sp species-level matrix was then converted to a 5 × 1,000,000 specimen-level matrix using data in the third row (species abundance), with each column corresponding to an individual specimen in the parent population (N = 1,000,000).

Model rock volumes were populated by randomly drawing specimens, without replacement, from the parent distribution and then randomly assigning them locations in the rock volume (Figs. 2, 3B). The generic bivalve and gastropod shells created using the Raup model served as the basis for generating specimens of different species that varied in both size and shape. These generic model shells possess shell height, width, and length values of 1 mm. Therefore, in order to generate a specimen of a particular size, the generic set of coordinates were linearly scaled so that the specimen had a shell height equal to the size bin value assigned to its species. For simplicity, we chose to associate each species with only a single size value and ignored intraspecific size variation. Specimen shape was modified by the model in association with size modification. While the shell heights were scaled only by the size bin value, the shell lengths and widths were scaled additionally by the length or width aspect ratios, respectively. Again, for simplicity, each species was associated with only a single shape and intraspecific shape variation was ignored.

Model realism demands that the model should not allow bivalve and gastropod shells to occupy the same space and thus crosscut one another. It was therefore necessary for the code to identify instances of shell crosscutting as the volume was being populated with specimens and correct for this overlap. While the volume was being populated, the code compared each newly added specimen to each previously added specimen, in the order of their addition to the volume, to assess whether specimens overlapped in space. If the range of coordinates along the x, y, and z axes of one specimen overlapped with that of any specimen previously placed in the volume, then the two specimens were considered to overlap. If the code detected overlap with any existing specimens, then the new-coming specimen was randomly assigned a new location in the volume and overlap was reassessed. This process was repeated until there was no overlap, at which point the code would move on to placing the next specimen. The number of specimens placed in a volume is defined by the user. For all model runs presented here, the number of specimens per populated volume was set equal to 50. In order to handle situations where no more specimens could be placed in a volume without creating overlaps, the code monitored the number of failed attempts to place a specimen. If this number went above 10,000, the code would terminate attempts to place specimens, display a ‘timed out’ message, and set the output variable ‘wntthru’ equal to 0. Aborted volumes were not used in analyses.

Once the volume was populated by specimens, up to three positions along the vertical axis between the top and bottom of the volume were selected at random (Fig. 3C). These points represent fracture horizons or bedding surfaces. All specimens that intersected one or more of these horizons were counted by the code, and their species identity and associated information recorded. These specimens constituted the fossil sample collected from that volume of rock (Fig. 3D). The code recorded information about each intersected specimen, including its species identity and size, and recorded this information in a master list of sampled specimens. The code continued to generate and sample populated volumes until a user-supplied sample size quota was met. This quota represents the sampling intensity and is represented in the horizontal axis in Figures 5–7. In different model runs, the quotas were set at 10, 50, 100, 150, 200, 250, and 300 specimens, allowing us to examine the effects of different sampling intensities on the magnitude of bias induced by preferential intersection.

For all but the smallest quotas, the number of specimens sampled from any one populated volume was below the quota. To reach this quota, the code therefore repeated the cycle of populating and sampling a new volume, each time adding newly intersected specimens to the master list. This cycle was repeated until the total number of specimens on the master list was equal to the user-defined quota. It is important to understand the relationship among the total number of splits, the per-volume splitting intensity, and the sample size quota. The total number of splits is inherently related to the sample size quota, as any time a greater number of populated volumes were generated as demanded by the sample size quota, there would also be an increase in the total number of splits. The number of splits per populated volume, however, was set by the user to be equal to 3 for all model runs.

Because a splitting plane can intersect multiple specimens, the code must also deal with circumstances where the addition of the final set of intersected specimens overshoots the sampling quota. In such circumstances, the code assessed whether the sample size quota had been met following each sampling of a populated volume. If, after the addition of the newly intersected specimens, the total number of specimens on the master list was not greater than the quota, then the entire set of newly intersected specimens was added. However, if the intersected specimens in the final splitting were more than needed to reach the quota, then the code randomly selected only the number of specimens needed to meet the quota.

Once the quota was reached, the code calculated the species richness, evenness, and average specimen size of the sample. This was done using the information on species identity and size that was included in the entry for each intersected specimen in the master list. Calculation of species richness was straightforward and the code simply tallied the number of distinct species in the master list. For evenness, the code calculated Pielou’s J (Pielou 1966), again based on the master list. For average specimen size, the code simply took the average size of all specimens on the master list regardless of species identity.

The procedure outlined above produced only a single set of values of species richness, evenness, and average specimen size for the generated bulk sample. This procedure was therefore repeated 150 times, and mean values and 95% confidence intervals for species richness, evenness, and mean specimen size were calculated based on the range of values yielded over the 150 model runs. This is straightforward for species richness and assemblage evenness. However, average specimen size merits further clarification. The code calculated the mean value of 150 average specimen size values, or the mean of means. The mean values and confidence intervals of species richness, evenness, and average specimen size were calculated for each of the seven sample quotas examined in the model (Figs. 5–7).

As a null model to compare against the model of preferential intersection, we also wrote code to generate a random draw model. This model simply draws a quota of specimens at random from the same parent distribution used to populate the model volumes. This model is analogous to fossil collection by sieving unconsolidated sediments and retrieving all specimens contained within a given volume of sediments. A random draw model has previously been used to model fossil collection via sieving in a study of biases associated with choice of mesh size (Kowalewski and Hoffmeister 2003). Like the preferential intersection model, the random draw model was repeated 150 times for each quota size in order to calculate the mean and 95% confidence interval of species richness, evenness, and average specimen size.

The nine parent populations described above provided a basis for assessing how the magnitude of bias induced by preferential intersection varied in response to changes in the species abundance distribution and the distribution of species among size classes. Each parent population was run through both the preferential intersection and random draw models. The resulting curves of mean values and confidence intervals for species richness, assemblage evenness, and average specimen size allowed us to examine how the magnitude of the bias changed with sample size for each combination of species abundance and size distributions. We also performed two-tailed t-tests on sets of 150 values produced by the preferential intersection and random draw models, at a sample size of 300 specimens, to determine whether the two models produced significantly different richness, evenness, and average specimen size. T-tests were performed using the ‘t.test’ function in R (R Core Team, 2014).

**Appendix II. Splitting effort and the biasing effects of preferential intersection**

In order to investigate the impact of per-volume splitting intensity on the magnitude of bias, we ran the preferential intersection model with the Roy et al. (2000) species size distribution and the lognormal species abundance distribution (μ = 1.2281, σ = 1.2408, and total species richness = 50), with the quota fixed at 100 specimens. We ran the model with per-volume splitting intensity at 1, 5, 10, 15, 20, and 25. Figure 10 shows how the magnitude of the bias induced by preferential intersection varied as the per-volume splitting intensity increased. When each volume was split only once, the preferential intersection model at a sample size of 100 specimens yielded a mean species richness of 24 species, 6 less than the mean of 30 species generated by the random draw model. This was improved to 26 species when each volume was split 10 times and reached 29 when each volume was split 25 times. A permutation test was performed using the permTS function in the R package ‘perm’ to test whether the distribution of richness values was significantly different between per-volume intensity of 1 and 10. The permutation test yielded a p-value of 6.88 E–16, strongly supporting the hypothesis that these two sets of species richness values represented distinct distributions. These results demonstrate that, for this combination of species size distribution and species abundance distribution, bias is not overcome until per-volume splitting intensity reaches approximately 10 slices. If we consider that the thickness of the volume is set to 10 cm, this is analogous to splitting fossiliferous rock with sufficient intensity that, on average, resulting split pieces are no thicker than 1 cm. This result indicates that any effects of preferential intersection could be overcome through a deliberate splitting strategy. As splitting intensity increases and splitting planes are more closely placed, small specimens are more likely to be intersected by at least one splitting plane. This demonstrates that sufficiently exhaustive sample preparation can overcome the effects of preferential intersection.