



**Thank you for downloading this document from the RMIT Research Repository.**

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.

RMIT Research Repository: <http://researchbank.rmit.edu.au/>

**Citation:**

Drumm, D and Greentree, A 2017, 'Microscopy as a statistical, Renyi-Ulam, half-lie game: a new heuristic search strategy to accelerate imaging', Scientific Reports, vol. 7, 14652, pp. 1-13.

**See this record in the RMIT Research Repository at:**

<https://researchbank.rmit.edu.au/view/rmit:45928>

**Version:** Published Version

**Copyright Statement:**

© The Author(s) 2017. Creative Common Attribution 4.0 International License.

**Link to Published Version:**

<https://dx.doi.org/10.1038/s41598-017-14876-x>

**PLEASE DO NOT REMOVE THIS PAGE**

# SCIENTIFIC REPORTS



OPEN

## Microscopy as a statistical, Rényi-Ulam, half-lie game: a new heuristic search strategy to accelerate imaging

Daniel W. Drumm  & Andrew D. Greentree 

Finding a fluorescent target in a biological environment is a common and pressing microscopy problem. This task is formally analogous to the canonical search problem. In ideal (noise-free, truthful) search problems, the well-known binary search is optimal. The case of half-lies, where one of two responses to a search query may be deceptive, introduces a richer, Rényi-Ulam problem and is particularly relevant to practical microscopy. We analyse microscopy in the contexts of Rényi-Ulam games and half-lies, developing a new family of heuristics. We show the cost of insisting on verification by positive result in search algorithms; for the zero-half-lie case bisectioning with verification incurs a 50% penalty in the average number of queries required. The optimal partitioning of search spaces directly following verification in the presence of random half-lies is determined. Trisectioning with verification is shown to be the most efficient heuristic of the family in a majority of cases.

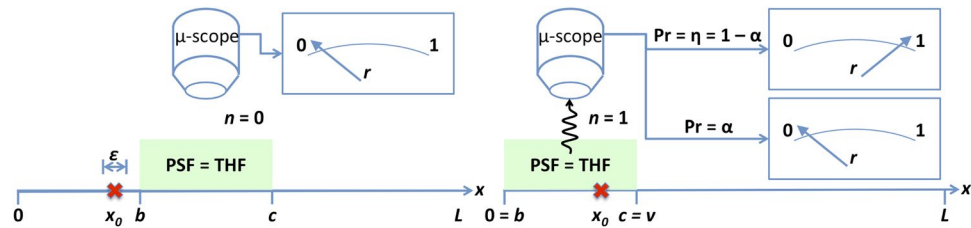
Optical microscopy remains a key platform technology enabling detection, tracking, and sometimes quantification of biologically and medically relevant targets. Standard and advanced microscopy approaches include confocal<sup>1</sup>, and multiphoton<sup>2</sup> techniques, which deliver high-power light to classically small (diffraction-limited) spot volumes. The probe light excites either endogenous targets<sup>3</sup>, or introduced emitters (suitably functionalised to bind targets<sup>4,5</sup> within or between cells).

Standard approaches to confocal and multiphoton microscopy involve rastering the focal spot through a sample at constant scan rate; this can be treated as a finite dwell time on each pixel/voxel. Firstly, this is to allow enough time for signal generation, and secondly, the dwell time is usually above a greater threshold value long enough that even dim regions of the sample exhibit low noise. However, biological materials (and some introduced emitters) often exhibit photosensitive or even phototoxic responses<sup>6</sup>, and may photobleach over time<sup>5,7,8</sup>. There is therefore a tension between resolving the image and minimally affecting the sample, and consequently a need to use the probe microscope light in the most efficient manner possible, extracting maximal information per photon used in both excitation and collection.

Whether using multiphoton or confocal techniques, a measurement illuminates a continuous sample region and is equivalent to asking a question of presence or absence of some target. However, both confocal and multiphoton microscopes suffer considerable signal photon loss<sup>1</sup>, due to their collection angles, detection point-spread functions, and sundry loss pathways. Conversely, due to improvements in detector technology, false photon gain events (*e.g.* dark counts) are now extremely rare. Therefore the physical situation strongly resembles a half-lie scenario, where the presence of a particle may be subject to a lie (*i.e.*, be undetected), but the absence of a particle is correctly reported—albeit by an absence of signal.

Half-lies have been studied in the context of games where a Questioner attempts to guess which integer the Responder has in mind of those in some domain (*e.g.*,  $[1, 10^6]$ ). Only queries requiring yes/no answers are allowed; hence the response space is binary, and the category of problem is known as binary search<sup>9,10</sup>. The optimal classical method for the Questioner to accomplish this aim is to consecutively halve that domain; this technique is a special case of the splitting algorithm<sup>11</sup>. For binary searches,  $\lceil \log_2(m) \rceil$  queries suffice to locate one of  $m$

Australian Research Council Centre of Excellence for Nanoscale BioPhotonics, Physics, School of Science, RMIT University, Melbourne, VIC, 3000, Australia. Correspondence and requests for materials should be addressed to D.W.D. (email: [daniel.drumm@rmit.edu.au](mailto:daniel.drumm@rmit.edu.au))



**Figure 1.** Microscopy as a game: determining a fluorescing target’s position (red cross)  $x_0$  on  $x \in [0, L]$  within  $\varepsilon$ . Top: microscope ( $\mu$ -scope) top-hat point-spread function (green-shaded area) not aligned with target ( $x_0 < b < c$ ); no photon generated ( $n = 0$ ); measurement result guaranteed  $r = 0$ . Bottom: target within microscope PSF ( $b < x_0 < c$ ); photon generated ( $n = 1$ ). Possibility of photon loss  $\alpha$ , leading to probabilistic measurement result ( $r \in \{0, 1\}$ ). Here,  $c = v$  as discussed in the main text; also  $v$  and  $\alpha$  are uniquely specified in each measurement. The game ends when the distribution for  $x$  is bounded within any distance  $\varepsilon$  not necessarily centred on the true location  $x_0$ , which is never found.

integers. The game becomes more interesting when the Responder occasionally lies<sup>12,13</sup>. These Rényi-Ulam games have been studied extensively, both mathematically<sup>14–18</sup>, and in various applications<sup>19–29</sup>.

As originally described, the Rényi game is of the combinatorial form “Find a number  $a \in \{1, 2, \dots, m\}$ ” using arbitrary queries with yes/no answers, where the Responder will lie a set number of times<sup>12</sup>. Ulam’s game is similar, but specifies the number of times the Responder may lie (without enforcing such behaviour). Variants since considered include asymmetric errors (including half-lies), asking only comparison-type questions, *etc.* (14 gives a comprehensive list). Partially fidelitous solutions have been presented for a subset of these problems<sup>30</sup>, though most studies to date have focussed upon finding optimal strategies where a minimum number of queries suffice in all cases to identify the target with 100% fidelity<sup>15–17,20,21,24–26,31–40</sup>.

An interesting variant of Rényi-Ulam games arises when the lies are limited to only one response type, *e.g.*, “yes”, or “present”. Such constraints are termed “half-lies”<sup>31</sup>, and have again been the subject of some enquiry<sup>21,36</sup>, albeit for specific known numbers of such half-lies, or for unfettered choice of query<sup>18</sup>. Another interesting subset of the games are those where the lies are no longer limited in number, but instead occur probabilistically<sup>12,30,41</sup> (otherwise known as the statistical variant). We are interested in the intersection of these two sub-types.

Here, we show the connection between microscopy and Rényi-Ulam games, and explore subtle differences from standard Rényi-Ulam games arising from the physics of confocal and multiphoton apparatus. These differences preclude the usual approach of determining an optimal number of queries, instead forcing us to consider their average number across many trials. We find a useful heuristic for accelerated searching under these conditions, which include the treatment of noise as an intrinsic characteristic of the measurement apparatus, rather than as a background distractor as assumed by signal detection theory<sup>42,43</sup>.

This paper is organised as follows: in Sec. 1 we describe the equivalence between Rényi-Ulam games and microscopy, before developing a mathematical approach to selecting queries in Sec. 2 and obtaining an information map. We discuss potential uses of the map in various strategies in Sec. 3, analysing their average behaviour across many trials, before concluding with Sec. 4.

## Microscopy as a Rényi-Ulam game

We wish to determine one (single-photon-emitting) fluorescent target’s position  $x_0$  in some continuous 1D domain  $x \in [0, L]$  to within some precision  $\varepsilon$  (or alternatively, to better than some threshold resolution  $\varepsilon^{-1}$ ). The precision may be considerably larger than the target’s physical size, which is always on the atomic to few-nanometre scale. The average number of queries required should be minimised, and the result achieved must be 100% fidelitous. We assume the microscope illuminates and collects from a region defined by a single top-hat function (THF) with arbitrarily controllable boundaries, as this is the simplest approximation we can make. This is equivalent to a query in the Rényi-Ulam game over the THF. Queries provoke  $n$  fluorescence signal photons from the target where  $n \in \{0, 1\}$ .  $n = 1$  iff the target is in the non-zero region of the THF, otherwise  $n = 0$ . These queries are formally equivalent to Rényi-Ulam interval queries, *e.g.*, “Is  $b < x_0 < c$ ?”, for some threshold values  $b$  and  $c$  (see Fig. 1). (Note, however, that under some circumstances—as described in Sec. 3.1—these are also equivalent to comparison queries.)

Microscopes often exhibit photon loss mechanisms, including finite numerical aperture, attenuation through optical fibres, and detector inefficiencies. Such system-wide losses are usually (multiplicatively) characterised through an overall efficiency,  $\eta$ . This parameter describes the chance of collecting a photon if the target is illuminated by the peak value of the point-spread function, and since photons are usually collected through the same PSF, if collection occurs through the peak value also. Most microscopes have non-flat PSFs, often modelled with Gaussian functions. Thus, the true efficiency is  $\eta_0 = \eta \times \text{PSF}^2(x_0)$ , if we normalise the peak value to 1. For simplicity, we treat the PSF as a flat top-hat, and hence the true efficiency is simply  $\eta$  (if the target is illuminated). Being concerned primarily with photon loss, we define it as the complement of collection events,  $\alpha_0 = 1 - \eta$ , regardless of the origin of the loss.

Photon loss thus occurs with independent probability  $\alpha_0 \in [0, 1]$ . This loss is the genesis mechanism of the half-lie. Surviving photons are collected and counted; their number is the reported measurement result  $r \in \{0, 1\}$ . We formally connect target presence with a Rényi-Ulam “yes”, its absence with a “no”, and allow half-lies on “yes”

System	State	Possible responses
Rényi-Ulam	No	No
	Yes	No or Yes
Microscopy	Absent	$r = 0$
	Present	$r \in \{0, 1\}$

**Table 1.** Possible responses in Rényi-Ulam game and microscopy experiment, illustrating the connection between the experiment and the game. Because the “No” and  $r = 0$  responses appear for both states, the possibility of a half-lie must be considered when one of these responses is obtained. Conversely, a “Yes” or  $r = 1$  response is guaranteed to be fidelitous.

states only. *I.e.*, a “yes” could be reported as a “no”, but a “no” state can never be reported as a “yes” (see Table 1, or Eq. 1). The game is immediately extensible in multiple commuting spatial dimensions; one plays a separable, independent game in each such dimension.

Since we have probabilistic, random photon loss, we explicitly play a statistical, rather than a combinatorial, Rényi-Ulam game. The number of loss events is theoretically unbounded. Therefore, optimal solutions to combinatorial Rényi-Ulam games as described above do not exist here, since they describe scenarios where the number of lies (or half-lies) is known *a priori*.

For cases where up to a certain total fraction of the queries may be lied to (in any order, and in adversarial fashion), it has been shown<sup>44</sup> that an  $O(\log_2 n)$  questioner’s solution exists for lie-rates  $< 1/3$ , whilst for rates above the responder can always win. The authors note that their result is also obtainable from the proofs of Rivest, Meyer, Kleitman & Winklmann<sup>31</sup>. The complexity of the half-lie problem has been generally equated to that of the full-lie problem, for specified numbers of half-lies<sup>31</sup>, and the number of necessary and sufficient queries for one half-lie has been solved<sup>36</sup>. However, this still does not describe the probabilistic nature of the half-lies inherent to microscopy. The combination of half-lies and the statistical variant does not appear to have been yet considered.

Since microscopes typically have the probabilistic  $\alpha_0 > 1/3$ , but do not operate in adversarial fashion, we instead concern ourselves with minimising the *average* number of queries. Such an approach will aid experiments requiring many replications and/or variations, as are common in biology and medicine.

### Mathematical formalism

We now formulate the game, with particular regard for the choice of measurement boundary (defined shortly) and the consequences of that choice for algorithm run time. Random variables  $X$ ,  $N$ , and  $R$ , are defined on the domains of  $x$ ,  $n$ ,  $r$  (position, number of signal photons, measurement result);  $A$  is the signal loss probability, on  $\alpha \in [0, 1]$ , and  $V$  is the turn-off boundary between the measured and non-measured regions, defined on  $v \in [0, 1]$ . Without loss of generality due to the invariance of Shannon information with a reordering of the distribution, we fix the turn-on boundary at the leftmost point with PDF  $> 0$  for convenience. This also further specifies the query type as a comparison question.

The physics of our system allows us to put constraints on their behaviour:

$$\begin{aligned}\mathbb{P}_{N|V,X}(n|v, x) &= \theta(v - x) \delta_{n,1} + \theta(x - v) \delta_{n,0}, \text{ and} \\ \mathbb{P}_{R|N,A}(r|n, \alpha) &= \delta_{n,0} \delta_{r,0} + \delta_{n,1} [\alpha \delta_{r,0} + (1 - \alpha) \delta_{r,1}]\end{aligned}\quad (2)$$

where  $\theta$  is the Heaviside step function, and  $\delta_{ij}$  is the Kronecker delta function as the domains  $n$  and  $r$ , of  $N$  and  $R$  respectively, are discrete. Since we only probe emitters left of  $v$ , we can only excite emissions left of  $v$ ; the receipt of a photon necessarily requires its emission, and is sometimes blocked by signal loss. For conciseness, we henceforth omit the probabilities’ domains.

As the domain  $v$  is continuous, we set  $\mathbb{P}_V = \delta(v - v_0)$ , using the Dirac delta function, and assuming a perfectly known boundary,  $v_0$ . Also,  $\mathbb{P}_A = \delta(\alpha - \alpha_0)$ , a perfectly known false negative rate,  $\alpha_0$ . Further, although the target has a unique position  $x_0$ , this is unknown; therefore we have a uniform prior over  $X$ , such that  $\mathbb{P}_X = 1/L$ .

Since  $A$  and  $V$  are uniquely determined,

$$\begin{aligned}\mathbb{P}_{A|N,V,X} &= \mathbb{P}_A; \\ \mathbb{P}_{V|N,X} &= \mathbb{P}_V.\end{aligned}\quad (2)$$

$$\begin{aligned}\text{Also, } \mathbb{P}_{N,V,X} &= \mathbb{P}_{N|V,X} \mathbb{P}_V \mathbb{P}_X, \\ &= \mathbb{P}_{N|V,X} \mathbb{P}_V \mathbb{P}_X, \text{ and} \\ \mathbb{P}_{N,X} &= \int_V \mathbb{P}_{N|V,X} \mathbb{P}_V \mathbb{P}_X dv.\end{aligned}\quad (3)$$

We now consider that when we attempt to detect photons, we are making a quantum measurement of the field. Since the position of the emitter and measurement boundary are determined before the Markov process of measuring (collapsing) the state and reporting the outcome, we can deconditionalise thus:

$$\mathbb{P}_{R|N,A,V,X} = \mathbb{P}_{R|N,A}.\quad (4)$$

We form the full joint by:

$$\begin{aligned} \mathbb{P}_{N,A,V,X,R} &= \mathbb{P}_{R|N,A,V,X} \mathbb{P}_{N,A,V,X} \\ &= \mathbb{P}_{R|A,V,N,X} \mathbb{P}_{A,V|N,X} \mathbb{P}_{N,X} \\ &= \mathbb{P}_{R|A,V,N,X} \mathbb{P}_A |V,N,X \mathbb{P}_{V|N,X} \mathbb{P}_{N,X} \end{aligned} \tag{5}$$

Substituting Eqs 1–4 in Eq. 5 and marginalising gives

$$\begin{aligned} \mathbb{P}_{X,R} &= \sum_N \int_{A,V} \mathbb{P}_{N,A,V,X,R} d\alpha dv \\ &= \frac{1}{L} [\delta_{r,0} \theta(x - v_0) + \alpha_0 \delta_{r,0} \theta(v_0 - x) \\ &\quad + (1 - \alpha_0) \delta_{r,1} \theta(v_0 - x)], \text{ and} \\ \mathbb{P}_R &= \int_X \mathbb{P}_{X,R} dx \\ &= \frac{1}{L} [\delta_{r,0} (L - v_0 + \alpha_0 v_0) + \delta_{r,1} (1 - \alpha_0) v_0] \end{aligned} \tag{6}$$

Now, using the definitions of Shannon information:

$$\begin{aligned} H(X) &= - \int_X \mathbb{P}_X \log_e (\mathbb{P}_X) dx, \\ \text{and } H(X|R) &= \sum_R \int_X \mathbb{P}_{X,R} \log_e \left( \frac{\mathbb{P}_R}{\mathbb{P}_{X,R}} \right) dx, \end{aligned} \tag{7}$$

we can calculate the mutual information:

$$\begin{aligned} I(X: R) &= H(X) - H(X|R) \\ &= \log_e \left( \frac{1}{L} \right) - \frac{1}{L} [(1 - \alpha_0) v_0 \log_e (v_0) \\ &\quad + (L - v_0 + \alpha_0 v_0) \log_e (L - v_0 + \alpha_0 v_0) \\ &\quad - \alpha_0 v_0 \log_e (\alpha_0)] \end{aligned} \tag{8}$$

which we maximise to find the condition:

$$v_0 = \frac{L \alpha_0^{\frac{\alpha_0}{1-\alpha_0}}}{1 + (1 - \alpha_0) \alpha_0^{\frac{\alpha_0}{1-\alpha_0}}}, \tag{9}$$

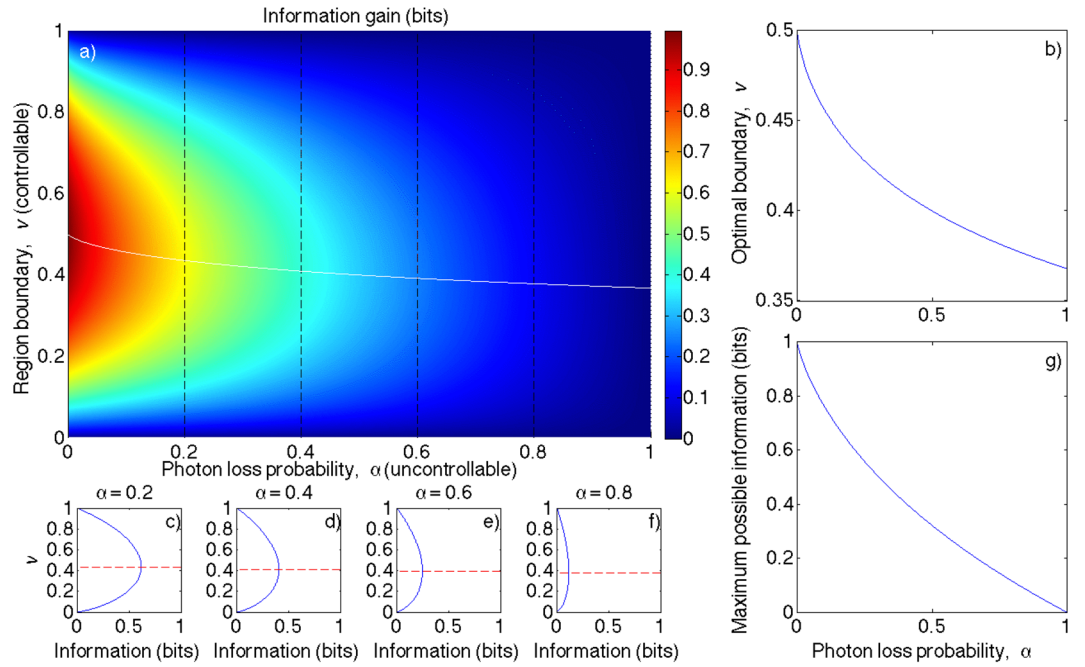
shown in Fig. 2. As expected, since this limit is formally equivalent to the canonical binary search problem,  $\lim_{\alpha_0 \rightarrow 0} v_0 = 1/2$ .  $v_0$  is undefined for  $\alpha_0 = 1$ ; however, the limiting value as  $\alpha_0 \rightarrow 1$  is  $v_0 \rightarrow 1/e$ . The optimal measurement domain in the presence of photon loss (half-lies) is therefore less than  $L/2$ .

### Search strategies

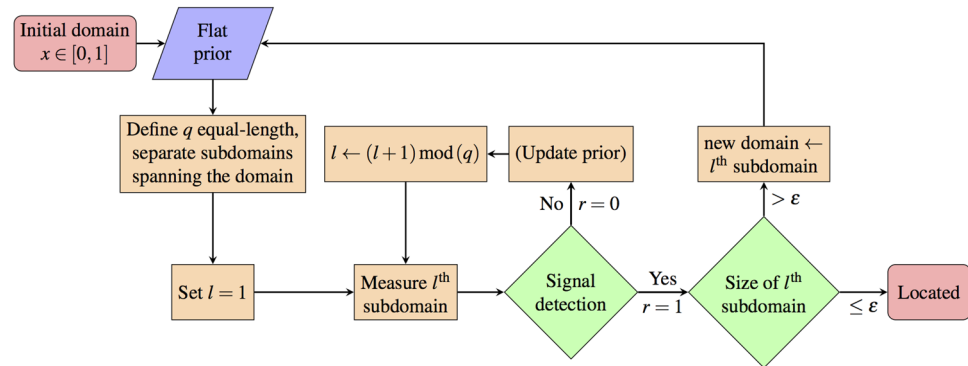
**Proposals.** A greedy strategy for target searches would be to maximise the information gain for each subsequent measurement. However, any search over a region that is not a simple dividend of the overall region will be difficult to update for a null result,  $r = 0$ . (It would also return the query type to interval questions.) We therefore consider simple heuristic approaches, whereby the domain is split into  $q$  subdomains which are explored sequentially until the target is explicitly located, whence the split (into  $q$  subdomains) recurs. Note that the final split in each heuristic is still into  $q$  subdomains, even if fewer would suffice to meet the criterion of precision  $\epsilon$ , maintaining the heuristics’ simplicity. We develop and contrast four such approaches: bi-, tri-, and tetra-sectioning with verification, and the limiting extension of these to sequentially scanning each of the  $\epsilon^{-1}$  subdomains.

The obvious choice for an heuristic is bisectioning, the limiting case for zero half-lie rate. (Basic bisectioning is equivalent to binary search in this limit.) Here, we define an approach where null measurements simply unbalance the PDF, and following measurements are undertaken on the other half; *i.e.*, naïvely rastering between only two equally sized subdomains at a time. This cycle repeats until positive verification of target presence occurs by a measurement  $r = 1$ ; the problem is then recursively reposed within the successful half (Fig. 3, for  $q = 2$ ). We call this process “bisectioning with verification”. Note that by insisting on positive verification, we cannot take advantage of all the available information in the low-half-lie limit. On average, this process dictates 1.5 measurements per level of enquiry, due to the equal probabilities of the target being or not being in the first-examined subdomain; *i.e.*, for precision  $\epsilon$ ,  $1.5 \lceil \log_2 (\epsilon^{-1}) \rceil$  queries are required.

Here, we extend the definition of bisectioning with verification to define a new family of heuristics:  $q$ -sectioning with verification.  $q$  equal subdomains are queried in turn until positive verification is obtained. The average number of measurements will be  $\bar{m}_{q,\epsilon,\alpha_0}$ ; for half-lie-free search this is:



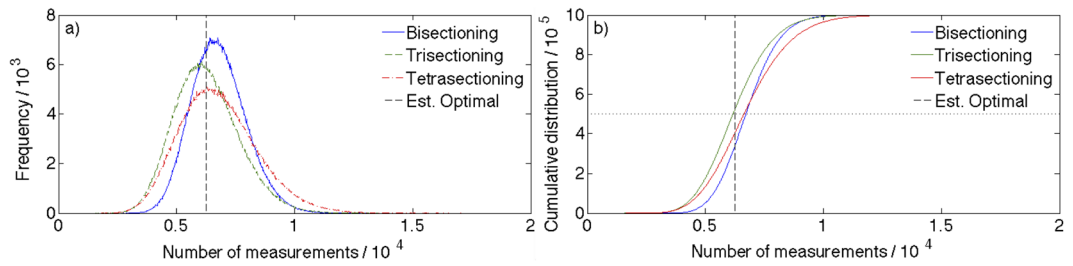
**Figure 2.** Optimal measurement domain strategy per independent measurement on  $[0, L]$  in the presence of photon loss half-lies: (a) false-colour plot of information gained as a function of false negative rate  $\alpha$  and region boundary  $v$ , with particular values of  $\alpha$  marked (black dashed lines), and optimal region boundary placement shown as a function of  $\alpha$  (white curve); (b) focus on optimal boundary placement (white curve from (a)); (c–f) line slices of (a) showing the information gain for each boundary placement at  $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$  (black lines in (a)); and (g) maximal information possible per measurement shown as a function of  $\alpha$ .



**Figure 3.**  $q$ -sectioning algorithm requiring detection to proceed. Update prior step not formally required. We choose (without loss of generality) to index subdomains from the left boundary.  $l$  is an index for the  $q$  subdomains required at any step of  $q$ -sectioning. The algorithm completes when the final subdomain is smaller than the precision required,  $\epsilon$ .

$$\bar{m}_{q,\epsilon,0} = \frac{(q + 1)}{2} \lceil \log_q(\epsilon^{-1}) \rceil. \tag{10}$$

Now we consider trisectioning with verification; we have  $\bar{m}_{3,\epsilon,0} = 2 \lceil \log_3(\epsilon^{-1}) \rceil$ . Due to our high probability of signal loss, we expect  $\bar{m}_{3,\epsilon^{-1},\alpha_0 \neq 0}$  to be a considerably higher number of measurements than  $\bar{m}_{3,\epsilon^{-1},0}$ . Note that this is explicitly not the ternary search of Rényi-Ulam games, defined as a special case of  $q$ -ary search<sup>34,45</sup>. That refers instead to a similar, but subtly different process in which the Responder indicates which of the  $q$  subdomains holds the target. A good example problem of that type is finding a heavier coin by balance weighing<sup>18</sup>. Neither is it the ternary search of computer science or learning, where it is used to maximise unimodal functions by evaluating their value at two intermediate points<sup>46,47</sup>, or to classify and sort data<sup>48</sup>.



**Figure 4.** Results of 1,000,000 Monte Carlo trials with  $\alpha_0 = 0.99$  for each sectioning strategy: (a) Histograms of search heuristic results indicating the mode of each strategy, and estimated optimal average result, and (b) cumulative distributions of same, indicating median values. Note: the trisectioning trace is to the left of both bi- and tetra-sectioning for almost the entire number of trials. Also note the estimated optimal average (dotted black line), similar but not equal to the trisectioning median (intersections with dashed black line).

We also extend this idea to tetrasectioning, which is analogous to bi- and tri-sectioning but partitions the surviving search space into four equal subdomains at any level of measurement. Similarly, the average number of measurements for half-lie-free search will be  $\bar{m}_{4,\varepsilon^{-1},0} = 2.5[\log_4(\varepsilon^{-1})]$ .

The penultimate approach we consider is to use precisely as many equal sections as we require for the desired precision, *i.e.*,  $\varepsilon^{-1}$  subdomains, and to raster over these subdomains until positive verification is obtained. We call this approach “rastering with verification”. On average, this will require  $\bar{m}_{q=\varepsilon^{-1},\varepsilon^{-1},0} = 0.5(\varepsilon^{-1} + 1)\log_{\varepsilon^{-1}}(\varepsilon^{-1})$  measurements, which for any appreciable value of  $\varepsilon^{-1}$  is prohibitively large compared to the previous three strategies. We therefore summarily dismiss rastering with verification as unfeasible (except for small  $\varepsilon$ , where it approaches the other strategies) and do not consider it further in this work.

Finally, we consider rastering as is commonly performed. Here, the domain is continuously scanned from left to right, with the THF having width  $\varepsilon$ . This is used instead of dwelling on each subdomain in turn to avoid ring-down of the equipment after stopping and any consequent dark time to allow its mitigation. The scan rate is slow for several reasons, including shot noise suppression, and convenience in automation by standardising the process for every pixel. We can, however, model this approach as physically dwelling on each subdomain in turn for a set number of queries before moving on. For a dwell time corresponding to  $\gamma$  queries, this approach will require  $\gamma/\varepsilon$  measurements;  $0.5\gamma/\varepsilon$  on average if the scan is adaptively implemented and will cease after locating the target. Not only does this approach again require a far larger number of measurements for any appreciable  $\varepsilon^{-1}$ , but the inefficiency is compounded by the dwell parameter,  $\gamma$ , which must be at least 1. Once more, we dismiss this strategy and ignore it henceforth.

We note in each of these heuristic schemes that since the initial domain PDF is uniform, the entropy of an initial query on an edge subdomain is the same as one on an internal subdomain (of equivalent length). Therefore, the query types, though often of interval form, are informationally equivalent to comparison queries. Similar arguments can be made for the following measurements until a new mapping within a subdomain occurs and the reasoning recurs.

**First-pass analysis.** To estimate the average optimal number of queries needed to locate a target with precision  $\varepsilon$  for specific  $\alpha_0$ , we consider the amount of mutual information – from Eq. 8 – delivered by a hypothetical optimal first measurement as dictated by Eq. 9. We approximate all consequent measurements with this value. Achieving the desired precision corresponds to acquiring  $\log_2(\varepsilon^{-1})$  bits of information, and thus requires an estimated  $\lceil [\log_2(\varepsilon^{-1}) \text{ bits}] / [I(X: R)|_{\alpha_0, \nu_0} \text{ bits/query}] \rceil$  queries.

We simulated one million random emitter positions, using  $\alpha_0 = 0.99$ , and subjected each to bi-, tri-, and tetra-sectioning with verification as described above, to an arbitrarily selected precision of  $\varepsilon = 10^{-10}$ . (This high precision displays the algorithmic speedup well; microscopy generally deals with  $\varepsilon > 10^{-6}$ , or the approximate ratio of a confocal spot diameter to the length of a 96-well plate.) The measurement regions were curtailed only after positive detection, otherwise cycling through the 2, 3, or 4 sections available. For  $\alpha_0 = 0.99$ , this dictates an estimated  $\approx 6240$  “optimal” queries.

Figure 4a shows a histogram of the Monte Carlo results: the mode  $[Mo()]$  of each stratagem is apparent, with  $Mo(\text{trisectioning}) < Mo(\text{tetrasectioning}) < Mo(\text{bisectioning})$ . The trisectioning mode is also smaller than the estimated optimal average number of measurements; the semi-infinite nature of the data space allows for an average value pulled right of the mode.

Figure 4b is the cumulative density function for each approach, from which the median is extractable where each curve crosses the horizontal dotted line. Again, the same ordering is evident, with  $\text{Median}(\text{trisectioning})$  falling just below the estimated optimal average number.

The distribution means are displayed in the first row of Table 2, and are consistent with the other measures of central tendency, except that the trisectioning mean is now slightly higher than the estimated optimal average. In addition, the mean number of queries required was evaluated for several different half-lie rates, and form the remainder of Table 2. Similar estimates for the optimal behaviour based on  $I(X: R)|_{\alpha_0, w}$  are also provided.

The requirement of absolute knowledge (zero PDF outside the final location), by positive verification of presence through signal capture, inflicts a penalty on the sectioning strategies. As mentioned above, 33.2 bits of information are required: at zero half-lie rate, this dictates 34 measurements in a perfect scheme. The final row

$\alpha_0$	Bi-sectioning	Tri-sectioning	Tetra-sectioning	Est. optimal
0.99	6783.13	6275.92	6771.67	6240
0.5	119.00	105.02	110.49	104
0.3	80.15	69.00	71.64	66
0.1	58.55	49.01	50.06	44
1/18	54.99	45.71	46.49	40
1/35	53.00	43.85	44.60	37
1/36	52.94	43.80	44.44	37
0.01	51.69	42.64	43.18	35
0.001	51.07	42.06	42.57	34
0	51.00	42.00	42.50	34

**Table 2.** Average number of measurements (for  $10^6$  trials) to locate an emitter to one of  $10^{10}$  subdomains using various heuristic search strategies and an estimate of the optimal average value obtained from Eqs 8 and 9 (shown in Fig. 4g).

of Table 2 shows that none of the sectioning strategies performs particularly well compared to this mark, with bisectioning using on average 50% more queries than required. In contrast, tri- and tetra-sectioning only require 24% more queries than are optimal. Neither uses the extra (positive) information which can be derived from null results, which faithfully indicate that the target is not present. The ignorance and disuse of the full information available by the insistence on positive verification translates as an extra cost in the number of measurements required.

It is instructive to consider the lower (non-trivial) limit of half-lie rates, where the most likely outcome is one half-lie. Assuming the 34 necessary plus one erroneous measurements, one half-lie is most likely for  $1/36 \leq \alpha_0 \leq 1/18$ , with the expected value of one half-lie for  $\alpha_0 = 1/35$ . Rivest *et al.*<sup>31</sup> proved that finding a target in  $k$  discrete subdomains, when up to  $E$  responses to comparison questions may be erroneous *iff* the truth is of one specified type (e.g., less than), requires  $Q \geq \lceil \log_2(k) + E \log_2 \log_2(k) + O[E \log_2(E)] \rceil$  comparison questions. For our example case, where  $E = 1$  and  $k = 10^{10}$ ,  $E \log_2(E) = 0$ , requiring  $Q \geq \lceil \sim 38.2 \rceil = 39$  queries. From Table 2, the trisectioning heuristic performs on average within 11% of the Rivest *et al.* bound within these limits, and outperforms the bi- and tetra-sectioning approaches.

Not only does bisectioning fail to be optimal (even for zero half-lies), trisectioning outperforms it at every half-lie rate as well (for  $\epsilon^{-1} = 10^{10}$ ). From Table 2 we can determine that the penalty for using bi- instead of tri-sectioning is 8–21%, depending on the photon loss rate—and worst for low loss rate (where the insistence on positive verification is most deleterious). There is therefore an immediate, quantifiable benefit to preferring the trisectioning strategy over bisectioning. Tetrasectioning has a similar, albeit smaller, penalty, and is nearly as efficient as trisectioning for low half-lie rates.

Although for low photon loss, the trisectioning heuristic requires more measurements than the estimated optimal information use, its performance improves markedly as losses increase. For typical confocal microscope conditions ( $\alpha_0 \approx 0.99$ ), trisectioning requires on average only 1% more measurements, and beats bisectioning by 10%.

Returning to the original context of confocal/multiphoton microscopy, one major point of difference is that here, we explicitly recognise that once a subdomain is searched and a null result obtained, the target is more likely elsewhere in the full domain. In contrast, conventional approaches dwell on each pixel, effectively performing a number of measurements (as defined by our approach) several orders of magnitude larger than the one performed here before moving the PSF. Such processes are information inefficient.

The second, and more impactful, difference is that conventional microscopy *always* searches the same amount of area per query, whereas the sectioning approaches resize the PSF after verification. This is the origin of the change from linear to logarithmic behaviour—if the local intensity of the probe beam within the PSF can be maintained for diffuse measurements.

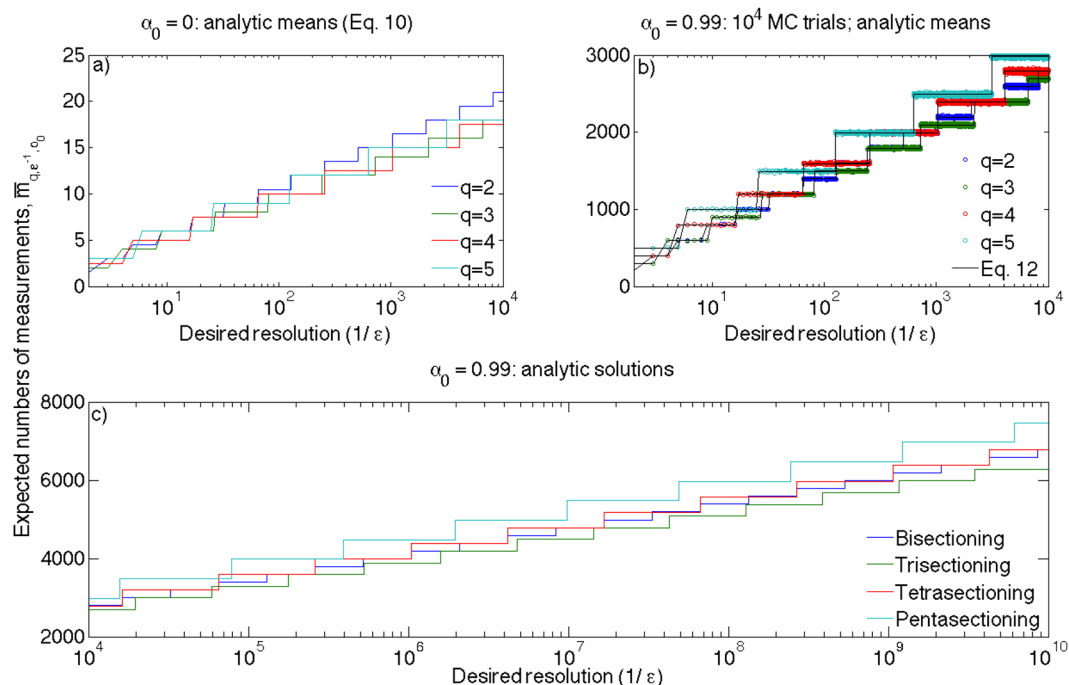
**Exploration for other precisions. Zero half-lie rate.** The selection above of  $\epsilon^{-1} = 10^{10}$  was somewhat arbitrary, and it is natural to question whether the performance of the trisectioning algorithm was due to the nearness of  $10^{10}$  to a power of three (and distance from powers of two and four).

To explore this concept, we propose the following thought experiment: let us set the half-lie rate to  $\alpha_0 = 0$ . Now, there is no probabilistic component to the average number of searches required to find a target in  $q$  subdomains; the average number is simply  $\frac{q}{2}$ , based on the random placement of the target within the initial domain. We can therefore easily compute the expected number of searches under a given heuristic as  $\frac{q}{2} \lceil \log_q(\epsilon^{-1}) \rceil$ . Performing this for  $q \in \{2, 3, 4, 5\}$  and  $\epsilon^{-1} \in \{2, 3, 4, \dots, 10^6\}$ , we compute Fig. 5a.

It is quickly apparent that trisectioning and tetrasectioning dominate a search for minimum-search strategies. Further investigation shows that tetrasectioning is actually the most frequent winner, for 611612 of the considered values of  $\epsilon^{-1}$ . It is unique for 569970 of these cases. See Table 3 for more detail.

Strategies with  $q > 8$  are never efficient, hence our comments about rastering as generally performed. Strategies with  $5 \leq q \leq 8$  are efficient only in limited special cases. Choosing  $q \in \{3, 4\}$  offers optimal behaviour for 987,353 of the cases considered.





**Figure 5.** Average numbers of searches required to find a target for various  $\epsilon$ , given particular heuristics and  $\alpha_0$ . (a) Directly calculated exact means for  $\alpha_0 = 0$ ; (b) Mean results of running  $10^4$  Monte Carlo trials for each value of  $\epsilon^{-1} \in \{2, 3, \dots, 10^4\}$  and  $\alpha_0 = 0.99$ ; (c) direct calculation of expected means for the same conditions as (b); (d) extension of direct calculation of (c) for all  $\epsilon^{-1} \in \{10^4 + 1, 10^4 + 2, \dots, 10^8\}$ .

q	Solution type	
	Optimal	Unique
2	2	1
3	415,229	246,610
4	611,612	569,970
5	142,433	12,634
6	1,500	11
7	11	1
8	1	0
10	0	0
100	0	0
1,000	0	0

**Table 3.** Performance of various  $q$ -sectioning search heuristics for  $1/\epsilon \in \{2, 3, \dots, 10^6\}$  and  $\alpha_0 = 0$ . Performance is described as optimal where no  $q$ -sectioning heuristic solution exists for that  $\epsilon^{-1}$  with fewer measurements, and unique if no similarly optimal solution for that  $\epsilon^{-1}$  exists with different  $q$ .

*Realistic microscopy parameters.* The situation becomes more interesting for non-zero  $\alpha_0$ , and yet harder to obtain extensive data for. Since non-zero  $\alpha_0$  implies some probabilistic behaviour, we again performed many Monte Carlo simulations to obtain the equivalent of one datum from Fig. 5a, but for  $\alpha_0 = 0.99$ . We chose  $\epsilon^{-1} = 10^6$  as equivalent to a search across a microscope slide or well plate (of order 10 cm) down to the diffraction limit of light (of order 100 nm). It is also conveniently substantially closer to powers of 2 and 4 than 3, as given by our metric function for the effective nearness of  $\epsilon^{-1}$  to a power of  $q$ :

$$F(q, \epsilon) = 0.5 - \left\| 0.5 - \left[ \log_q(\epsilon^{-1}) - \left\lfloor \log_q(\epsilon^{-1}) \right\rfloor \right] \right\| \quad (11)$$

A value of  $F = 0$  means that  $\epsilon^{-1}$  is an exact match to a power of  $q$ , whilst  $F = 0.5$  is the greatest distance attainable.

We find that trisectioning is still competitive (Table 4). We have also included pentasectioning for completeness; it performs poorly compared to the others, despite  $10^6$  being a similar distance from a power of five as from a power of three. Bi- and tetra-sectioning, which have similar  $F$  values, have drastically different modes. Clearly, the relative nearness to a power of  $q$  is not the only contributing factor to the performance of the heuristics.

$q$	$F(q, \varepsilon)$	Mean	Median	Mode
2	0.068	3990.2	3924	4157
3	0.425	3887.4*	3789*	3674
4	0.034*	3983.5	3853	3458*
5	0.416	4482.8	4318	3657

**Table 4.** Measures of central tendency from results of Monte Carlo trials in the same vein as Fig. 4;  $10^6$  trials for  $\alpha_0 = 0.99$ . The lowest values in each column are denoted by asterisks (\*).

$\varepsilon^{-1}$	$q$	$F(q, \varepsilon)$	Mean	Median	Mode
$10^3$	2	0.034	1994.7	1929	1813
	3	0.288	2092.7	1994	1762
	4	0.017*	1992.0*	1861*	1568*
	5	0.292	2489.5	2325	2016
$3 \times 10^3$	2	0.449	2394.6	2328	2108
	3	0.288	2391.7	2292	2158
	4	0.225	2389.7*	2257*	1890
	5	0.025*	2429.5	2329	1834*
$10^4$	2	0.288	2797.7	2716	2641
	3	0.384	2690.3*	2591*	2607
	4	0.356	2787.8	2656	2359*
	5	0.277*	2989.1	2823	2628
$3 \times 10^4$	2	0.127*	2992.8	2927*	2821
	3	0.384	2988.7*	2980	2602
	4	0.436	3188.9	3059	2565*
	5	0.405	3845.9	3321	3330
$10^5$	2	0.390	3391.9	3326	3145
	3	0.480	3288.8*	3190*	2947*
	4	0.305	3584.9	3452	3417
	5	0.153*	3983.7	3820	3704
$3 \times 10^5$	2	0.195	3790.7	3725	3657
	3	0.480	3587.0*	3488*	3071
	4	0.097*	3985.2	3852	3637
	5	0.164	3984.9	3823	3160
$10^6$	2	0.068	3990.2	3924	4157
	3	0.425	3887.4*	3789*	3674
	4	0.034*	3983.5	3853	3458*
	5	0.416	4482.8	4318	3657

**Table 5.** Measures of central tendency from results of Monte Carlo trials in the same vein as Fig. 4;  $10^6$  trials for  $\alpha_0 = 0.99$ . The lowest values in each column/major row are denoted with asterisks (\*).

We explored a limited subset of  $\varepsilon^{-1}$  to the same rigour (Table 5). Here, pentasectioning occasionally performs comparably to the other heuristics, but mostly considerably poorer than the best of them. It is notable that the heuristic with minimal distance is most often not the one with the lowest mean number of measurements.

**General solution.** It being impractical to perform  $10^6$  simulations for each value of  $\varepsilon^{-1}$ , we match Fig. 5a by lowering our standards. We performed  $10^4$  Monte Carlo simulations for every value of  $\varepsilon^{-1}$  between 2 and  $10^4$ , with  $\alpha_0$  set to 0.99; the results can be seen in Fig. 5b. We note that the individual heuristics still perform in step-wise semilogarithmic fashion (albeit with larger jumps); however, the pattern is significantly distorted from the zero-half-lie case. Our observations regarding pentasectioning are borne out; in the main it performs the worst of these heuristics. We also note the enhanced performance of the trisectioning heuristic over the tetrasectioning, as evidenced by its greater proportion of being the minimum solution—especially at the right of the figure, where the logarithmic scale greatly compresses the data over  $\varepsilon^{-1}$ .

The regularity of each heuristic's behaviour led us to propose a more general form for the mean number of measurements, based upon the frequentist interpretation of probabilities as the mean chance of outcome given exhaustively many trials:

$q$	2	3	4	5
Optimal	1,687	14,253,472	2,522,055	1
Unique	1,687	14,253,472	2,522,055	1

**Table 6.** Performance of various  $q$ -sectioning search heuristics for  $1/\varepsilon \in \{2, 3, \dots, 2^{24}\}$  and  $\alpha_0 = 0.99$ . Performance is described as per Table 3: optimal where no  $q$ -sectioning heuristic solution exists for that  $\varepsilon^{-1}$  with fewer measurements, and unique if no similarly optimal solution for that  $\varepsilon^{-1}$  exists with different  $q$ .

$\varepsilon^{-1}$	Type	$q$			
		2	3	4	5
$10^3$	Estimate	1994.7	2092.7	1992.0*	2489.5
	Analytic	1995	2093	1992.5*	2490
$3 \times 10^3$	Estimate	2394.6	2391.7	2389.7*	2429.5
	Analytic	2394	2392	2391*	2490
$10^4$	Estimate	2797.7	2690.3*	2787.8	2989.1
	Analytic	2793	2691*	2789.5	2988
$3 \times 10^4$	Estimate	2992.8	2988.7*	3188.9	3845.9
	Analytic	2992.5	2990*	3188	3486
$10^5$	Estimate	3391.9	3288.8*	3584.9	3983.7
	Analytic	3391.5	3289*	3586.5	3984
$3 \times 10^5$	Estimate	3790.7	3587.0*	3985.2	3984.9
	Analytic	3790.5	3588*	3985	3984
$10^6$	Estimate	3990.2	3887.4*	3983.5	4482.8
	Analytic	3990	3887*	3985	4482

**Table 7.** Comparison of means, estimated from  $10^6$  Monte Carlo trials, versus those calculated directly from Eq. 12. The lowest values in each column are denoted with asterisks (\*).

$$\bar{m}_{q,\varepsilon^{-1},\alpha_0} = \left[ q \left( \frac{1}{1-\alpha_0} - 1 \right) + \frac{q+1}{2} \right] \lceil \log_q(\varepsilon^{-1}) \rceil. \quad (12)$$

The first term in the prefactor recognises that in the long run,  $1/(1-\alpha_0)$  trials must occur in the correct subdomain before a photon will be detected. However, for each trial (except the final one), we incur the penalty of searching the other  $q-1$  subdomains, necessitating  $q \left( \frac{1}{1-\alpha_0} - 1 \right)$  searches per level before we expect to receive a signal on the next pass through. On the final pass (per level), we have  $(q+1)/2$  searches as before. This total, of course, is then scaled by the logarithmic ceiling function, which dictates how many levels of search need be accomplished to sufficiently locate the target. Of course,  $\lim_{\alpha_0 \rightarrow 0} \bar{m}_{q,\varepsilon^{-1},\alpha_0} = \bar{m}_{q,\varepsilon^{-1},0}$  from Eq. 10.

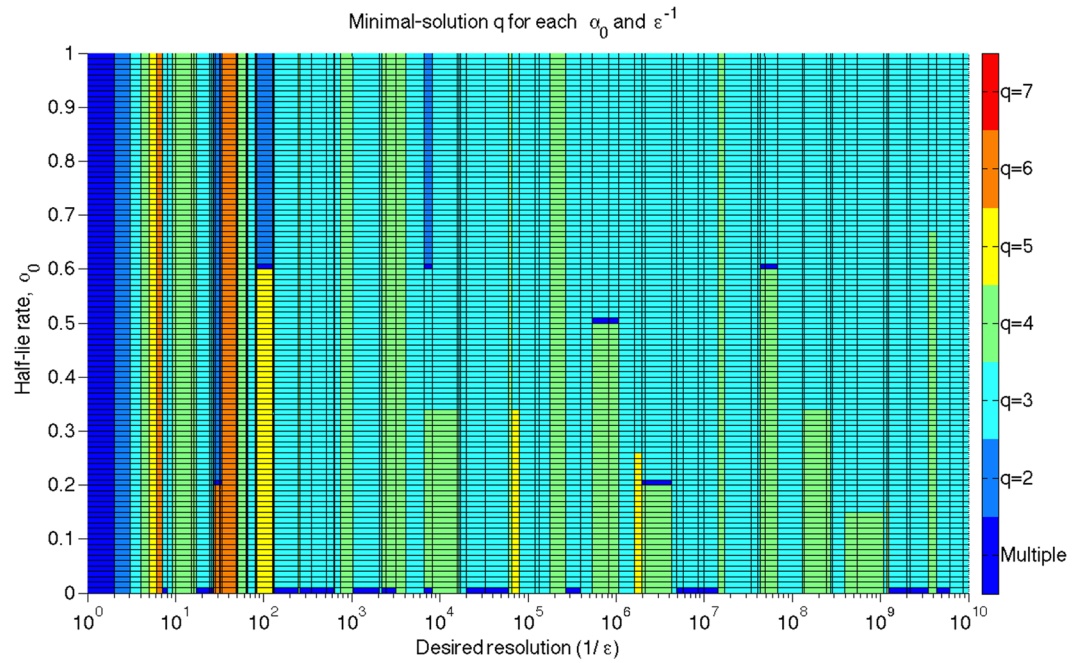
Figure 5b also presents a direct calculation (by Eq. 12) of the means previously estimated through Monte Carlo sampling. By eye, it appears to describe the behaviour well. We extend its domain over more values of  $\varepsilon^{-1}$  than it is practical to simulate, even for a single  $\alpha_0$ , in Fig. 5c. Given the semi-logarithmic nature of the plot, it is clear that trisectioning search is the optimal  $q$ -sectioning heuristic for the vast majority of  $\varepsilon^{-1} \in \{2, 3, \dots, 10^{10}\}$  (and  $\alpha_0 = 0.99$ ). Closer exploration reveals it to be the optimal  $q$ -sectioning heuristic for all  $\varepsilon^{-1} > 2^{24} \sim 1.7 \times 10^7$  (and  $\alpha_0 = 0.99$ ). It is also optimal for 85.0% of the cases below  $2^{24}$ , as Table 6 shows. What is less clear, given the scale of the axes, is that *all* of the solutions are now unique. This is due to the more complex prefactors involved.

It is now trivial to compare the estimates of the means from Table 5 with Eq. 12 – see Table 7. The agreement is very good, with most estimates within 1 measurement of the corresponding analytic values.

**Behaviour with changes in loss.** Obtaining a general analytic form allows us to explore behaviour of the family of  $q$ -sectioning with verification heuristics across changes not only to the desired precision, but also of the half-lie rate  $\alpha_0$ . Due to the step-like nature of the  $\bar{m}_{q,\varepsilon^{-1},\alpha_0}$  function, its behaviour can be modelled simply by taking values on each side of the step function locations—which are the powers of the various  $q$ . Figure 6 shows the optimal heuristic type(s) for each  $\varepsilon^{-1} \in \{2, 3, \dots, 10^{10}\}$  and  $\alpha_0 \in \{0, 0.01, \dots, 1\}$ .

The abundance of light blue in Fig. 6 shows that the trisectioning with verification heuristic is optimal not only for most  $\varepsilon$  at  $\alpha_0 = 0.99$ , but also at lower half-lie rates. This is a somewhat surprising result; above  $\varepsilon^{-1} = 128$ , trisectioning dominates the plot – except for zero half-lies. Of course, the  $q$ -sectioning with verification heuristic is not optimal at zero half-lies, since it does not take advantage of the information contained in failures to detect photons; basic bisectioning (binary search) search is faster there. However, searches not requiring verification cannot be 100% fidelitous for any  $\alpha_0 > 0$ .

As we might expect, the next most successful heuristic is tetrisectioning with verification. Bisectioning makes a few brief appearances, but is not uniquely optimal for any  $\alpha_0$  after  $\varepsilon = 2^{13}$ . Pentasectioning is similar, though rarer again, and is no longer uniquely optimal under any conditions after  $q = 5^{13}$ .



**Figure 6.** Minimal  $q$ -sectioning with verification heuristic;  $q: \text{Min}_q(\bar{m}_{q,\varepsilon^{-1},\alpha_0})$  for each  $\varepsilon^{-1} \in \{2, 3, \dots, 10^{10}\}$  and  $\alpha_0 \in \{0, 0.01, \dots, 0.99\}$ . The  $\varepsilon^{-1}$  grid is set at a unique sorted list of  $\varepsilon^{-1} \in \{q^i - 1, q^i, q^i + 1\}$  for  $q \in \{2, 3, 4, 5\}$  and for any  $i$  giving a value of  $\varepsilon^{-1}$  in the desired range. Panel colours show the minimal solution type; panels are shaded by the solution type(s) at their lower-left vertices, although due to the step-like nature of  $\bar{m}_{q,\varepsilon^{-1},\alpha_0}$  the solution types are valid for all  $\varepsilon^{-1}$  across each panel horizontally. Note: no solutions are given for  $\alpha_0 = 1$  as these are undefined.

In principle, Fig. 6 could be extended further in  $\varepsilon^{-1}$ ; however near  $2^{49}$ , double-precision floats fail to avoid numerical errors in calculating  $\bar{m}_{q,\varepsilon^{-1},\alpha_0}$ . Extending Fig. 6 this far would unacceptably compress the plot, which for clarity is limited to cases already discussed.

## Conclusions

In summary, the  $q$ -sectioning with verification family of search algorithms has been proposed and studied. Verification is an intriguing constraint on search algorithms; its cost is most apparent at low photon loss rates. For zero loss, bisectioning with verification ignores half of the available information and suffers a consequent loss of efficiency compared to the optimal binary search. However, for any non-zero loss, algorithms requiring verification retain 100% fidelity in their predictions, while  $q$ -ary searches cannot (except in the unattainable limit of infinite measurements).

An analytic formula describing the mean number of measurements for entire the family of  $q$ -sectioning with verification algorithms given any choice of  $q$ , precision  $\varepsilon$ , and half-lie (photon loss) rate  $\alpha_0$  has been developed. The family's behaviour has been mapped across a wide range of  $\varepsilon$  and  $\alpha_0$ , and the optimal choice is not simply a function of the relative nearness of  $\varepsilon$  to a power of  $q$ .

Trisectioning with verification is most often the optimal  $q$ -sectioning with verification heuristic for problems involving any random half-lies and precision  $\varepsilon < 1/128$  of the initial search domain. Further, although trisectioning with verification is sub-optimal, it appears near-optimal for our example case according to our estimator.

In practical terms, since photon-counting microscopes have already been developed, the trisectioning scheme discussed here offers the chance for greatly accelerated target searches over current confocal rastering techniques. This approach could be particularly useful for locating short-lived species if the microscope point-spread function can be modified quickly enough, via both changing the spot size as well as shifting the beam centre, thereby changing the search domain rapidly between measurements. However, before such a task is undertaken, it would be beneficial to reconsider this problem with explicit Gaussian roll-offs and/or general shape to the PSF, and also to fully characterise the intensity changes required by the varying PSF widths. Finally, before such an approach is applied to biological systems, attempts should be made to include penalty functions characterising photon exposure and/or damage, as well as accounting for practical considerations such as time to modify or shift the PSF, to study how the dynamics of an optimal biological search differ from those presented here. Information may not be the sole metric of interest to be optimised; a full cost-benefit analysis accounting for these issues may eschew the information benefit of shifting position more often.

## References

- Pawley, J. B. *Handbook of Biological Confocal Microscopy* (Springer, 2006).
- Horton, N. G. *et al.* In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nature Photonics* **7**, 205 (2013).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263**, 802 (1994).
- Reineck, P. & Gibson, B. C. Near-Infrared Fluorescent Nanomaterials for Bioimaging and Sensing. *Advanced Optical Materials* (2016).
- Reineck, P. *et al.* Brightness and Photostability of Emerging Red and Near-IR Fluorescent Nanomaterials for Bioimaging. *Advanced Optical Materials* **4**(10), 1549 (2016).
- Kochevar, I. E. Phototoxicity Mechanisms: Chlorpromazine Photosensitized Damage to DNA and Cell Membranes. *The Journal of Investigative Dermatology* **76**, 59 (1981).
- Song, L., Hennink, E. J., Young, I. T. & Tanke, H. J. Photobleaching Kinetics of Fluorescein in Quantitative Fluorescence Microscopy. *Biophysical Journal* **68**, 2588 (1995).
- Shaner, N. C. *et al.* Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nature Methods* **5**(6), 545 (2008).
- Nowak, R. Generalized binary search. *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on.* (IEEE, 2008).
- Nowak, R. Noisy generalized binary search. In *Advances in neural information processing systems*, p1366 (Curran Associates, Inc., 2009).
- Garey, M. R. & Graham, R. L. Performance Bounds on the Splitting Algorithm for Binary Testing. *Acta Informatica* **3**, 347 (1974).
- Rényi, A. On a problem of information theory. *MTA Mat. Kut. Int. Kozl.* **6B**, 505 (1961).
- Ulam, S. M. *Adventures of a Mathematician*, p. 281 (Scribner, New York, 1976).
- Pelc, A. Searching games with errors—fifty years of coping with liars. *Theoretical Computer Science* **270**, 71 (2002).
- Ellis, R. & Yan, C. Ulam's pathological liar game with one half-lie. *Int. J. Math. Math. Sci.* **2004**(09), 1523 (2004).
- Ellis, R. B., Ponomarenko, V. & Yan, C. H. The Rényi-Ulam pathological liar game with a fixed number of lies. *J. Comb. Th. A* **112**, 328 (2005).
- Ellis, R. B., Ponomarenko, V. & Yan, C. H. How to play the one-lie Rényi-Ulam game. *Disc. Math.* **308**, 5805 (2008).
- Cicalese, F. *Fault-Tolerant Search Algorithms: Reliable Computation with Unreliable Information* (Springer-Verlag, 2013).
- Yao, A. C. & Yao, F. F. On fault-tolerant networks for sorting. *SIAM J. Comput.* **14**, 120 (1985).
- Ravikumar, B., Ganesan, K. & Lakshmanan, K. B. On Selecting the Largest Element In Spite of Erroneous Information. *Annual Symposium on Theoretical Aspects of Computer Science* (Springer Berlin Heidelberg, 1987).
- Lakshmanan, K. B., Ravikumar, B. & Ganesan, K. Coping with erroneous information while sorting. *IEEE Trans. Comput.* **40**, 1081 (1991).
- Feige, U., Peleg, D., Raghavan, P. & Upfal, E. Computing with noisy information. *SIAM J. Comput.* **23**, 1001 (1994).
- De Bonis, A., Gargano, L. & Vaccaro, U. Group testing with unreliable tests. *Inform. Sci.* **96**, 1 (1997).
- Ngo, H. Q. & Du, D. Z. A Survey on Combinatorial Group Testing Algorithms with Applications to DNA Library Screening. In *Discrete Mathematical Problems with Medical Applications: DIMACS Workshop 55*, 171, Am. Math. Soc (2000).
- Cicalese, F. & Mundici, D. Learning and the Art of Fault-Tolerant Guesswork. In *Adaptivity and Learning: An Interdisciplinary Debate*, Eds Kühn, R., Menzel, W., Ratsch, U., Richter, M. M. & Stamatescu, I.-O. (Springer-Verlag, 2003).
- Mancini, S. & Maccone, L. Using Quantum Mechanics to Cope with Liars. *Int. J. Quant. Inf.* **3**(4), 729 (2005).
- Karp, R. M. & Kleinberg, R. Noisy binary search and its applications. In *Proc. 18th ACM-SIAM symposium on discrete algorithms*, Soc. Ind. and Appl. Math (2007).
- Corsi, E. A. & Montagna, F. The Rényi-Ulam games and many-valued logics. *Fuzzy Sets and Systems* **301**, 37 (2016).
- Jedynak, B., Frazier, P. I. & Sznitman, R. Twenty questions with noise: Bayes optimal policies for entropy loss. *J. Appl. Prob.* **49**(1), 114 (2011).
- Pelc, A. Searching with known error probability. *Th. Comp. Sci.* **63**, 185 (1989).
- Rivest, R. L., Meyer, A. R., Kleitman, D. J. & Winklmann, K. Coping with Errors in Binary Search Procedures. *J. Comp. Sys. Sci.* **20**, 396 (1980).
- Ravikumar, B. & Lakshmanan, K. B. Coping with known patterns of lies in a search game. *Theoretical Computer Science* **33**(1), 85 (1984).
- Pelc, A. Solution of Ulam's Problem on Searching with a Lie. *Journal of Computational Theory Series A* **44**, 129 (1987).
- Muthukrishnan, S. On optimal strategies for searching in the presence of errors. In *Proceedings of the Fifth ACM-SIAM symposium on Discrete algorithms*, p680 (1994).
- Dhagat, A., Gács, P. & Winkler, P. On playing "twenty questions" with a liar. In *Proc. 3rd ACM-SIAM symposium on discrete algorithms*, Soc. Ind. Appl. Math (1992).
- Cicalese, F. & Mundici, D. Optimal Coding with One Asymmetric Error: Below the Sphere Packing Bound. In *Computing and Combinatorics: Proc. 6th Int. Comp. Comb. Conf.*, Eds Du, D.-Z., Eades, P., Estivill-Castro, V., Lin, X. & Sharma, A. (Springer-Verlag, 2000).
- Cicalese, F. & Deppe, C. Quasi-Perfect Minimally Adaptive q-ary Search with Unreliable Tests. In *Proceedings of the 14th International Symposium on Algorithms and Computation* (Springer-Verlag, 2003).
- Cicalese, F., Deppe, C. & Mundici, D. Q-Ary Ulam-Rényi Game with Weighted Constrained Lies. In *Proceedings of the 10th International Computing and Combinatorics Conference* (Springer-Verlag Berlin, 2004).
- Cicalese, F. & Deppe, C. Perfect minimally adaptive q-ary search with unreliable tests. *Journal of Statistical Planning and Inference* **137**, 162 (2007).
- Xing, S. M., Liu, W. A. & Meng, K. Rényi-Berlekamp-Ulam searching game with bi-interval queries and two lies. *Discrete Applied Mathematics* **202**, 8 (2016).
- Schalkwijk, J. P. M. A Class of Simple and Optimal Strategies for Block Coding on the Binary Symmetric Channel with Noiseless Feedback. *IEEE Transactions on Information Theory* **IT-17**(3), 283 (1971).
- Peterson, W. W., Birdsall, T. G. & Fox, W. C. The Theory of Signal [sic] Detectability. *Transactions of the IRE professional group on information theory* **4**(4), 171 (1954).
- Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics* (Wiley, New York, 1966).
- Spencer, J. & Winkler, P. Three thresholds for a liar. *Combinatorics, Probability and Computing* **1**, 81 (1992).
- Soskov, I. N. Definability via Enumerations. *J. Symb. Logic* **54**(2), 428 (1989).
- Dobkin, D. P. & Souvaine, D. L. Detecting the intersection of convex objects in the plane. *Computer aided geometric design* **8**(3), 181 (1991).
- Salehin, K. M. & Rojas-Cessa, R. Combined methodology for measurement of available bandwidth and link capacity in wired packet networks. *IET communications* **4**(2), 240 (2010).
- Bentley, J. L. & Sedgewick, R. Fast Algorithms for Sorting and Searching Strings. *SODA* **97**, 360 (1997).

## Acknowledgements

The authors thank Nicolas Menicucci, Brant Gibson, Adrian Dyer, Jair Garcia, and Chris Xu for insightful discussions, and acknowledge the support of the Australian Research Council (project numbers CE140100003 and FT160100357).

## Author Contributions

D.W.D. and A.D.G. conceived of, planned, carried out, and analysed the study. D.W.D. and A.D.G. wrote the manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017