

This is a repository copy of *Building Confidence in Quantitative Systems Pharmacology Models : An Engineer's Guide to Exploring the Rationale in Model Design and Development.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/107642/>

Version: Accepted Version

Article:

Timmis, Jonathan Ian orcid.org/0000-0003-1055-0471, Alden, Kieran James, Andrews, Paul et al. (5 more authors) (2016) Building Confidence in Quantitative Systems Pharmacology Models : An Engineer's Guide to Exploring the Rationale in Model Design and Development. CPT Pharmacometrics Systems Pahrmacology. pp. 1-41.

<https://doi.org/10.1002/psp4.12157>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

¹ Building Confidence in Quantitative Systems
² Pharmacology Models: An Engineer's Guide to
³ Exploring the Rationale in Model Design and
⁴ Development

⁵ J. Timmis^{1,2,*%,}, K. Alden^{1,%}, P. Andrews², E. Clark², A. Nellis²,
B. Naylor^{1,2}, M. Coles ³, P. Kaye ³

⁶ ¹Department of Electronics, The University of York, UK.

⁷ ²SimOomics Limited, 106 Heworth Green York UK.

⁸ ³Centre for Immunology and Infection, Hull York Medical School/University of York,
⁹ York. UK.

¹⁰ *Corresponding author: jon.timmis@york.ac.uk

¹¹ %Authors contributed equally to this work

¹²

¹³

¹⁴

¹⁵ Word count introduction 1123

¹⁶ Word count body 5893

¹⁷

¹⁸ This tutorial promotes good practice for exploring the rationale of systems
¹⁹ pharmacology models. A safety systems engineering inspired notation ap-
²⁰ proach provides much needed rigour and transparency in development and

21 application of models for therapeutic discovery and design of intervention
22 strategies. Structured arguments over a models development, underpinning
23 biological knowledge, and analyses of model behaviours, are constructed to
24 determine the confidence that a model is fit for the purpose for which it will
25 be applied.

26 Introduction

27 When constructing a quantitative systems pharmacology (QSP) model there are many
28 issues to consider, from what aspects of the biological system needs to be modelled, hence
29 defining the scope of the model, to what modelling approach to use, through to how the
30 model is developed, and what abstractions are to be made during the model development
31 process. Likewise, there may be existing models that have been developed and are in
32 use as part of an experimental study, but which may be seen as a blackbox where the
33 rationale for their construction, use, and analysis is undocumented or was never coherently
34 established.

35 During model development various decisions have to be made, such as the inclusion of
36 simplifications and assumptions in place of biological knowledge, which may be very rea-
37 sonable but often are forgotten about or poorly documented. Yet these decisions impact
38 the relationship between any predictions that model generates and the real biological
39 system the model is aiming to capture, in turn impacting the level of confidence a re-
40 searcher has in applying those predictions within their own studies. Work in Alden et al
41 [4] presented a tool, Artoo, that permits the application of an adapted version of Goal
42 Structuring Notation (GSN) [30] through which a structured argument is developed to
43 show that a model is fit for the purpose for which it has been conceived. Within the
44 context of modelling, an argument is constructed by making claims concerning aspects of
45 model development, which are, where possible, supported by available evidence. In their
46 description, Alden et al provide an overview of using argumentation to examine fitness
47 for purpose, exemplifying application of the approach to explore the rationale underlying
48 the development of a previously published simulation of secondary lymphoid organ devel-
49 opment [7]. Thus Artoo was presented in a manner where claims were developed about a
50 specific model, rather than focusing on the process by which claims could be developed

51 and how different types of evidence can be used to establish those claims. Of critical
52 importance to that process, from which everything else flows, are two simple questions:
53 1) Has the right model been developed to address the specific question of interest? and
54 2) has the model been built correctly to address the specific question?

55 On the surface these might sound like obvious questions to ask and people might be
56 convinced that they have indeed satisfied both questions in a positive manner. However,
57 what is the evidence for such an assertion? If the model developer was asked to provide
58 clear evidence that their model is indeed fit for purpose, what evidence would be presented,
59 and how would that evidence be presented? Consider a number of issues associated with
60 model development: 1- what is the scope of your model in terms of the pharmacological
61 question you intend to ask? 2- who, or what, have you relied on for the underlying evidence
62 to build the model? 3 what assumptions did you make with respect to the biological
63 system you are working on and how it works? 4 what assumptions did you make when
64 moving from understanding your biological system, into mathematics? 5- why did you
65 choose a particular modelling style over another; and there are potentially many more
66 questions that could be asked. Indeed, alongside prompting these questions, adopting
67 such an approach can support inter-team working, having to explain, and document,
68 the rationale behind model development can promote greater transparency in the model
69 itself, and open it to wider scrutiny, which in the longer term, will promote better model
70 development.

71 These questions are routinely addressed in the area of safety engineering, where ensur-
72 ing that the correct device has been built, and that the device has been built correctly are
73 potentially of critical importance. Consider a simple example, the airplane. One assumes
74 there are some basic things to get right when building an airplane, for example the need
75 for wings and an engine, but what you build also depends on what the plane is to be
76 used for. Is it a transport plane or a passenger plane? Is it to be used for short distance,
77 or long distance? Ensuring you get the requirements clear ahead of time is important,
78 so understanding the purpose for which the plane is to be used is an essential part of
79 that process. Equally important is ensuring that what was required, was built correctly.
80 Were the right materials used?, was a rigorous engineering process undertaken?, was the
81 plane tested appropriately?, are there instructions on how to use it?, and have you taken
82 appropriate steps to identify and address possible sources of risk? Safety is now taken

83 for granted by passengers and we are, rightly, assured that safety is a primary concern
84 when building and using aircraft. Often that industry, and others, make use of safety
85 cases through the process of GSN to establish an argument for the safety of a system
86 [13, 15, 18].

87 Whilst developing and using a QSP model is not the same as building an aircraft,
88 there are analogies between the processes that leads to the construction and application
89 of both. A QSP model might be used as a key decision-making tool in determining dosing
90 regimes or within clinical trials [9, 10], which has potential safety critical implications, or
91 identify avenues of further (expensive) research that might otherwise be avoided. Whilst
92 we might not want to establish a safety case for a model, establishing that a model is
93 fit for the intended purpose for which it has been designed has the potential to increase
94 confidence, transparency and ultimate usage of such models in pharmacological studies
95 [20, 34]. GSN in the context of safety, and now in the context of model development has
96 been developed at York, and as yet is not widely used. However, it is through this tutorial
97 that it is hoped the wider use of such an approach will be adopted.

98 In this paper we provide a methodology that can be used to robustly develop ar-
99 gumentation structures that examine the rationale employed at various stages of the
100 development of a model. By encompassing all aspects of development, from composition
101 through implementation, analysis, and documentation; this approach provides a method-
102 ological structure with potential to increase confidence in the application of computational
103 models as predictive pharmacological tools. Although we ensure the focus is on the ar-
104 gumentation approach, we detail its application in the context of a mathematical model
105 of granuloma formation in the liver [3], a inflammatory immune response that occurs in
106 response to infection with the parasite Leishmania donovani. We show how exploring
107 the rationale behind the development of this simulation and assessing the composition of
108 the model after implementation, eases the assessment of simulation-derived predictions in
109 the context of the purpose for which this model has been designed: to explore potential
110 interventions that could further our understanding of treating this disease.

111 LEISHMANIASIS AND COMPUTATIONAL MOD- 112 ELS

113 Visceral leishmaniasis is a systemic tropical disease which, in the absence of treatment, is
114 usually fatal, with 20,000 - 40,000 deaths annually [8]. A defining feature of the immune
115 response to infection with *Leishmania donovani* parasites is the focal accumulation of
116 inflammatory cells within the liver: these aggregations are known as granulomas and pro-
117 vide a focus for immune mediated elimination of the parasite. The stages of the immune
118 response that follow infection and lead to granuloma formation and eventual parasite clear-
119 ance are illustrated in Figure S1. Importantly, the cellular composition of the granuloma
120 is dynamic and may comprise monocytes, T cells, and a range of other leukocytes includ-
121 ing B, NK, NKT, and dendritic cells in differing numbers and relative proportions [29].
122 Achieving an appropriate balance between cells that produce pro-inflammatory Th1-type
123 cytokines (e.g. IFN) and regulatory cytokines (e.g. IL-10) is important for stimulating
124 macrophages sufficiently to kill intracellular *Leishmania*, but without causing an over-
125 exuberant immune response that leads to destructive tissue pathology [29, 33]. Defining
126 how this balance across multiple cell types evolves over time during natural infection and
127 and how it might alter as a consequence of the administration of drugs and other therapies
128 provides a significant challenge in experimental immunology.

129 To generate insight into this important open question and move towards the devel-
130 opment of novel therapeutics against *Leishmania donovani*, experimental techniques are
131 required that are both less invasive and more ethically achievable than those used to study
132 HVL or EVL. Computational and mathematical approaches permit the development of
133 models that do not share the same constraints, and add capacity to interpret underlying
134 biological data [22] and to provide an experimental tool for exploring new hypotheses that
135 could be examined using traditional experimental approaches [11]. This methodology has
136 previously been employed in the development of a Petri net model of granulomatous in-
137 flammation in the liver of mice [3], motivated by the need to develop a tool capable of
138 generating insight into the importance of macrophage deactivation in immune regulation.
139 For the full design, implementation, and analysis detail that underlies this model we re-
140 fer the reader to the models accompanying publication and supporting materials [3]. To
141 provide a brief overview for the purposes of this tutorial, the Petri net [25] (notation in

142 Figure S2A)) captures biological entities involved in disease progression and resolution (T
143 cells, phagocytes, NKT cells, NK cells, and the Leishmania parasites) as places that hold
144 a number of counters. These counters signify the levels of each component at a particular
145 time-point of the simulation. Between each place are transitions that move tokens from
146 one place to another, decreasing or increasing the number of tokens as required (specified
147 by different line and arrow combinations, as shown in Figure S2(A). Each transition is
148 designed to capture a biological process, and is a mathematical construct controlled by
149 a number of parameters. At each timepoint the transitions between places fire at a rate
150 determined by probability density functions and the number of tokens in each place. The
151 simulation is designed to capture disease progression and resolution over an extended time
152 period. A high level overview of the leishmania Petri net model is reproduced from [3],
153 in Figure S2(B).

154 By running the Petri net model under different simulated physiological conditions (pa-
155 rameter exploration), the authors were able to suggest pathways through which regulation
156 of effector functions occur within the granuloma. Yet, for the potential of these insights
157 to further our understanding of the disease and impact therapeutic development to be
158 realised, it is vital that the composition, implementation, and analysis processes through
159 which the model has been developed are transparent and understood.

160 ENGINEERING TRANSPARENCY

161 In this section we outline a process using structured argumentation that assists the record-
162 ing of justifications and rationale for both the biological detail and engineering processes
163 that underlie the development of a computational model. The process and associated
164 tools to support that process take inspiration from the field of safety-critical systems,
165 where it must be demonstrated that a software system is as safe as reasonably practicable
166 [17]. Acceptable safety can be established and presented using arguments over evidence.
167 For increased accessibility and ease of communication, Goal Structuring Notation (GSN)
168 [30, 2] was developed as a visual notation for the presentation of arguments detailing
169 safety cases in critical systems engineering. The role of GSN in the wider safety commu-
170 nity is significant with various large industries making contributions to the GSN standard
171 [1].

172 In exemplifying an approach to expose the rationale underlying the development of
173 a model, we utilise, and suggest the use of, a previously published argumentation tool
174 by ourselves, Artoo [4], that permits the creation of a diagrammatic summary of the
175 structured argument of fitness for purpose. The semantics of the argumentation structure
176 employed in Artoo are inspired by that of GSN, with some modifications introduced
177 to allow an alteration of focus from safety cases to providing a rationale for fitness for
178 purpose. The argument is presented as a tree of connected argument components, of
179 specific shapes (Figure 1). The semantics are detailed in Figure S3. These components
180 start from a top-level claim (a GSN goal). At the beginning of the process a set of
181 fitness-for-purpose requirements (referred to as goals or claims, that the argument seeks
182 to substantiate) should be established, with an accompanying set of strategies that can be
183 used to assess whether the requirement has been met. The strategies typically break down
184 goals into sub-goals, and eventually link to evidence supporting the claim, alongside the
185 source of the evidence where appropriate. If a requirement cannot be fully supported by
186 available evidence, for example where there are gaps in the biological understanding, then
187 the assumptions and abstractions made in place of this evidence are documented, opening
188 all implementation decisions to scrutiny by other researchers in the field and identifying
189 areas of biological study that have been overlooked or require further laboratory work.
190 The process of constructing a claim using the semantics in Artoo is described in Figure
191 S5.

Figure 1: **Caption**

192 Arguing Fitness for Purpose

193 As outlined above, whereas GSN is applied to demonstrate evidence in safety cases, our
194 purpose is to develop a fitness for purpose argument with respect to a model. This change
195 in motivation introduces a subtle but important change to the semantics. When arguing
196 over safety, it is critical that a claim is terminated by a suitable evidence node supporting
197 that claim. However, when documenting our rationale that a model is fit for purpose,
198 the construction of an argument may not have a clear ending, in respect of there being
199 no available evidence to substantiate a claim [4]. Where this happens, this should not
200 automatically be seen as a weakness in the model, yet could instead reveal a number of

things. First, that a claim that is believed to be reasonable may in fact not be reasonable at all, and the process of constructing the argument has led to this conclusion. At this point, it might be wise to review the argument, alongside the model to investigate why this might be the case. Second, it might be that the claim is reasonable, but there is no evidence that is acceptable (as defined by the creator of the argument structure). In the case of arguing fitness for purpose, the claim can be left as undeveloped, that is the claim can remain in the argument structure, but highlights a clear gap in the evidence base, thus providing informative transparency of the lack of evidence to support the claim. Such a modification is vital in QSP modelling applications, where expert opinion and assumptions have to be used to mitigate the fact that the understanding of the biological system may be incomplete.

Taking the description in Figure S4 as a template of how to develop a claim, we turn attention to developing claims that encompass all stages encountered in model development. In Figure S5 we have split the process into seven distinct phases, all of which, we believe, greatly benefit from the adoption of a structured argumentation approach in revealing the rationale employed at that stage. To exemplify creation of argumentation at each phase, we now go through each in turn, providing case study examples in the context of leishmaniasis.

Step 1 - Define Purpose of the Model

As can be seen in Figure S5, understanding and defining the intended purpose of a model is a key part of the process, as the rationale for the other key phases of model development is strongly linked to that purpose. Purpose in this context can be defined as for what question the model is intended to answer. This purpose may vary from being a general model intended to explore a range of hypotheses and capture many components, or a very specific model that is intended for a distinct scientific question. In either case, a clear purpose should be defined and a clear scope of the model established, with key questions derived that the model will be used to address. The definition of the purpose forms the first stage in the construction of the argument structure: the top level claim. As described in Figure S3, this top level claim is usually associated with context nodes that define the key terms used to specify that purpose. From here, strategies are then set that will be used to argue that the top level claim is met: that the tool is fit for its specified purpose.

232 Figure 2 shows the top level of the argumentation structure used to explore the ratio-
233 nale underlying the development of the leishmaniasis model. The purpose of the model
234 is clearly stated: to explore the effects of the cytokine IL-10 on EVL, parasite infection
235 and regulation of granuloma formation. The top level claim is therefore made that the
236 model effectively captures EVL in the liver, thus a useful tool for meeting the intended
237 purpose. Attached to this claim are six strategies that will be used to support the claim.
238 It is hopefully easily noticed that these six claims correspond to the six rounded rectan-
239 gles in Figure 3: an examination of the rationale of each phase in the process of model
240 development. This section continues with examining each of these sections in turn.

Figure 2: **Caption**

241 **Step 2 - Assess available biological evidence**

242 Once a purpose has been defined, an understanding of the underlying pharmacological
243 and biological processes that will be used for the development of the model needs to be
244 established. It is often at this stage where the scope of the model can be compromised,
245 with the desire to include as much biological information as possible, but possibly at the
246 expense of simplicity (or necessity). Clear rationale for what biological and pharmacolog-
247 ical evidence is being used should be produced: without a specification of the data used or
248 any assumptions employed, it is difficult for researchers using model-derived predictions
249 to relate this prediction to their own experimental study. Step 2 of our process supported
250 by argumentation is used to assess (i) the scope of any supplied biological data; (ii) the
251 understanding gleaned from experts studying the biological system and (iii) the areas
252 of understanding that are currently lacking. For each of these, an argumentation claim
253 will be established and an appropriate strategy developed to support the claim. This all
254 contributes to creating the scope of the model. For example, evidence could exist as a
255 log of the experiment that collected the data, or a list of time-points at which the data
256 was collected. Employing this technique ensures that the model developer is aware of the
257 extent to which the current biological system is understood, and the scope of which any
258 data can be included in the developed model.

259 Figure 3 expands on the known Biology. At this stage of the process we are document-
260 ing what has been considered and collecting evidence for mechanisms and species without

making a judgement of whether they will be included in the model - this judgement is made in step 3. The strategy considers the cell populations, cytokines and chemokines that are mentioned in relevant literature. This is useful for generating a list of species that the modeller may later include, or exclude, depending on the weight of evidence for their involvement. Also on the top level is the micro-environment, which if correctly scoped, may exclude populations or mechanisms that fall outside the intended purpose of the model. As an exemplar for the purposes of this tutorial, we have expanded on the cytokines, showing a list of all the cytokines that are considered in the literature. Although the complete argument expands the rationale for inclusion of all cytokines, our exemplar expands on IL-10 and IL-1. For IL-10 it is thought that increasing levels of IL-10 are associated with parasite growth and suppresses parasite clearance [37, 19, 28]. IL-1 is a known pyrogen (meaning that it can cause the host body temperature to rise), and can potentially contribute to parasite killing through heatshock [31].

Figure 3: **Caption**

Step 3 - Rationale for Biological Assumptions

In step 2, consideration is given to the scope of the underlying biology and pharmacokinetics, without consideration of how this will be implemented in any model. However, that step may also have revealed areas of biological understanding that are incomplete, yet need to be included in the model. This can be seen in Figure 4, where the impact of the pyrogen IL-1 is noted as not being fully understood. Where such evidence gaps are identified, well informed, justified, assumptions will need to be introduced into the model. It is critical that the justification for any such assumptions are documented alongside the predictions generated by the model, as their introduction may have an influence on the validity of that prediction. If, for example, the purpose of the model is to produce predictions that inform laboratory research, it is vital that confidence in the assumptions are a fair reflection of the experimental system on which they will be testing this prediction: key when financial and technical resources have to be considered within a study.

In Figure 4 we expand on two examples from the cytokines that were being considered in Step 2. We demonstrate two common simplifying assumptions. For IL-1, the proposed mechanism of action on parasite load is killing of parasites indirectly via heat shock. It

290 can be argued that heat shock is neither necessary nor sufficient for parasite clearance,
291 as evidenced by the lack of impact of IL-1 receptor blockade on acquired resistance or
292 granuloma formation[33]. Considering the purpose of the model, it is reasonable to assume
293 that IL-1 can be excluded despite the fact that there is some evidence that it could impact
294 parasite load. This exclusion of IL-1 is one type of simplifying assumption. Figure 6
295 also shows a partially developed argument for merging IFN and TNF which ends in the
296 undeveloped claim that they perform the same function and can be merged into a single
297 proxy species. Both of these simplifying assumptions depend on the stated purpose of
298 the model for their potential validity. Both simplifying assumptions are to some extent
299 judgement calls that multiple stakeholders may wish to examine and influence, which
300 elucidates the importance of transparency and documentation of the argumentation.

Figure 4: **Caption**

301 **Step 4 - Rationale for Modelling Approach**

302 In implementing any model of a biological system, there may be several techniques that
303 could be selected (i.e. modelling paradigms, software tools). In this step, the model
304 developer can use argumentation to justify the engineering decisions taken during model
305 implementation. There can be a temptation to choose the modelling tool of convenience,
306 one that a modeller is familiar with, however, this can be a mistake. It is well known
307 that different modelling techniques can show different types of results and have an effect
308 on what is observed [14]. Therefore, it is important that the rationale for the choice of
309 modelling system be exposed. As an example, a claim could be made that an agent-based
310 modelling paradigm is most suitable for addressing the question of concern. Strategies
311 would then be employed to determine whether this is indeed the case, or whether other
312 approaches such as Ordinary Differential Equation (ODE) modelling would be more ap-
313 propriate. By using argumentation at this stage, the developer has a record of the im-
314 plementation decisions that were taken, with a fully evidenced justification of why these
315 decisions were taken.

316 Figure 5 shows a subsection of the argument concerning the modelling approach
317 adopted in the development of the Leishmaniasis simulation. From the top claim specified
318 in Figure 2, the strategy is to argue the appropriateness of the adopted approach, in this

case stochastic Petri nets. From here, our claim is that the adopted paradigm provides the means to represent the required aspects of the biological system. To support this claim, one would be required to compare the available approaches, and as such the stated strategies involve examining implementing the model as a Petri net, agent-based model, or ODEs. For the scope of this tutorial, Figure 5 expands on the Petri net suitability claim, arguing that we can capture the required stochasticity, capture granuloma heterogeneity, handle small integer number calculation, and produce an implementation that is computationally tractable. In this case we are able to evidence all four claims, suggesting we have a suitable approach for capturing the key aspects specified in the claim.

Figure 5: **Caption**

Step 5 - Rationale for Modelling Assumptions

By employing steps 2 and 3, any gaps in the biological understanding became apparent and were addressed via appropriately justified and documented assumptions. Previously we described how critical these assumptions were when relating the simulated system to the real system of interest. Additionally, this critical issue is also applicable when introducing simplifications that may be made during the development of the model. At this stage, it may be sensible to determine whether the full extent of the biological system of interest scoped in step 2 needs to be captured in the model. For example, modelling the impact of a number of cell receptors and their respective chemokines could potentially be reduced to a model of a single proxy chemokine and receptor pair, if what is being examined is the higher-level effect produced by these chemokines and receptors as an ensemble. An example of a similar issue could be a biological system consisting of tens of thousands of cells: complexity that may not be tractable to simulate. The simulation developer may determine that only capturing a percentage of that environment is enough to understand the overall emergent behaviour of that system. Taking a number of biological concepts and simplifying these into a single mechanism, or determining a biological concept to be unnecessary given the scope of the model, does however introduce assumptions that must both be taken into consideration when relating a model-derived result to the real system and be justified.

Figure 6 shows a subset of the argumentation structure produced from the top level

348 strategy to argue over the modelling abstractions. Similarly to previous examinations of
349 the biological information and assumptions, here, claims are made concerning the appro-
350 priate capture of the cells, cytokines, chemokines, and the environment. For the scope of
351 this tutorial, we have included the argument of one key assumption in the model: that
352 the dynamics of monocyte derived macrophages, dendritic cells, and neutrophils can be
353 adequately captured by a single cell type. Such an assumption reduces the complexity of
354 the model, yet could impact the meaning of any results generated. As such we support
355 this simplification with two claims: that parasites are not observed to replicate in these
356 cell types, yet these cells contribute to the cytokine microenvironment in the granuloma.
357 The first, supported by collaborators opinion, would suggest that these cells could po-
358 tentially be abstracted out of the model altogether, as they do not influence the models
359 purpose. However this is contradicted by the second, which makes the claim that these
360 cells contribute to the cytokine environment of interest. As such, we argue that these are
361 required, but can be abstracted to a single proxy cell type that expresses the cytokines
362 identified in Figure 3.

Figure 6: **Caption**

363 **Step 6 - Engineering the Implementation**

364 When going through this process alongside the development of a simulation, the developer
365 will now have justified the modelling approaches they are going to use (step 4) and the
366 abstractions they will make in implementation (step 5). The next step is to implement
367 the model. Issues of trust in simulations for science have previously been raised, and
368 much has been written on how this could be countered by the release of code [27, 26, 24].
369 However, we believe our approach to structured argumentation also provides a means of
370 increasing trust in the implementation alongside such arguments. Argumentation could,
371 for example, be used to argue that the code meets the specifications developed in the
372 previous phases above, and that an adequate testing routine has been developed and
373 performed.

374 Figure 7 shows a subset of such an argumentation structure for the Leishmaniasis
375 simulation, arguing that the system meets requirements for implementation and has been
376 adequately tested. The former is in some respects easier to show: claims can be made

377 concerning particular biological behaviours that are evidenced by aspects of the model
378 (such as equations), and links can be drawn to evidence derived on argumentation dia-
379 grams from previous phases of the process. Testing a complex simulation is much more
380 difficult. In Figure 7, the strategy to argue that the Leishmaniasis simulation was ade-
381 quately tested has been to ensure adequate structural coverage of the code by tests. In
382 this case, as is typically the case in high integrity software engineering, this strategy is
383 split into three phases: requirements testing [36]; unit testing [23]; and manual review.

384 Requirements testing ensures that the system has a collection of requirements describ-
385 ing the tasks that the system should perform, and it ensures that each requirement has an
386 associated test (or collection of tests) that demonstrates the system fulfilling the require-
387 ment. The requirements tests are run through the implementation to check that they pass
388 and to measure their structural code coverage. If all the requirements tests pass, then
389 this demonstrates that the implementation performs its tasks correctly. If all the require-
390 ments have appropriate tests that pass, then this demonstrates that the implementation
391 performs the correct tasks. If the requirements tests produce full code coverage, then this
392 demonstrates that the implementation performs only its tasks and nothing else.

393 In practice, it might be impractical to achieve full code coverage using just require-
394 ments tests at the system level. For example: there might be some error-checking code
395 deep within the call tree that is difficult to trigger under normal conditions. For these
396 cases, unit tests are used to inject particular values into the implementation to increase
397 the code coverage of the requirements tests.

398 Even using unit tests, it may not be possible to achieve full code coverage for some
399 types of code. For example: robustness checks, system libraries, or code that only executes
400 when running the system in a different mode. For these cases, the code is reviewed
401 manually to either determine that it will not execute in the situations we are providing,
402 or to argue why it does not need to be tested (for example, a commonly used system
403 library). Given the criticality of models we consider adequate testing to mean achieving
404 90% statement coverage and 90% branch coverage through requirements tests and unit
405 tests, with the remaining code reviewed manually.

Figure 7: **Caption**

406 **Step 7 - Justify experimental approach / analysis**

407 Once a simulation has been designed and implemented, model developers will perform
408 in silico experimentation and statistical analyses designed to elucidate biological insight
409 from the model [6]. However, for full transparency, the model developer should adopt an
410 argumentation approach to argue that the experiment is necessary and designed correctly,
411 prior to any simulation runs being performed. This will ensure that the time spent on
412 running complex simulations is minimised, and also ensure the analysis routines take into
413 account implementation inherent issues such as the inclusion of stochastic behaviours.
414 Results from the experiments and the analysis techniques employed to fully understand
415 the behaviour of a model need to be interpreted in terms of (i) the scope of the designed
416 simulation; (ii) the biological system being studied. The final stage of our process uses
417 evidence-based argumentation to draw conclusions from simulation-derived results, util-
418 ising the evidence compiled in Steps 1-5. Here, the simulation developer may make a
419 claim regarding some insight generated during the modelling project. They may then
420 draw on evidence from the complete argumentation process to show that the generated
421 insight can be supported. Figure 10 shows a subset of the argument that the experimen-
422 tal analyses performed are well designed and appropriate. This is divided into sub-claims
423 that describe two sets of experiments: (i) statistical analyses employed to understand the
424 behaviour of the model, and (ii) in silico experimentation used to perform experiments
425 that may be difficult to perform in the laboratory. Both sets of experiments are detailed
426 in [3]. Figure 8 shows one of each: appropriate sensitivity analyses for the first and IL-10
427 related experimentation for the second. In both cases the claims are supported by the
428 reasoning for the particular experiment, the experimental strategy, and the results. By
429 ensuring the design of such analyses is transparent, others using the result in their own
430 context are clear as to how each prediction has been derived.

Figure 8: **Caption**

431 USING ARGUMENTATION TO REFINE MODEL

432 PURPOSE, DESIGN and IMPLEMENTATION

433 Where the described process is employed, the key biological information to be modelled
434 will be identified, translated into a format that can be encapsulated within a computer
435 code, and developed into a computational model through which predictive experimenta-
436 tion can be performed. Completing a process where each step on this path is justified and
437 documented is advantageous in determining the degree to which predictions made can be
438 related to the real-world system being studied [4]. Whereas the process described above
439 focuses on that process of exploring the rationale of a model either during construction
440 or retrospectively, a completed argumentation structure should however not be seen as a
441 static document, and offers further advantages in cases where a model is to be repurposed
442 or refined.

443 As an example, consider the Leishmaniasis simulation that has been used as a case
444 study throughout this tutorial. This model captures the processes within EVL, an ex-
445 perimental mouse model of visceral leishmaniasis. However, the overriding objective is to
446 further our understanding of Leishmania in order to expedite the development of novel
447 therapeutics against the disease in humans. Although it is generally accepted that the
448 mouse provides an adequate model for exploring the disease in humans, this remains a
449 model of the disease in the mouse, and the links between this model and the human disease
450 need to be understood. One potential strategy could be to repurpose the model: altering
451 the focus to capture HVL rather than EVL. If this were to be undertaken, possessing a
452 rationale for the design, construction, and analysis of the computational model of EVL
453 would be very useful in determining the extent to which the model needs to be altered
454 to capture HVL. For example, an assessment of the biological information on which the
455 EVL model was constructed (step 2) and the assumptions that were introduced in that
456 model (step 3) would determine the relevance of that data to any model of HVL. Where
457 argumentation was used to construct the original model, we also argue that the approach
458 could be very useful in arguing over any alterations that are made if the purpose of the
459 model is adjusted.

460 Additionally, possessing a complete rationale detailing model development and analy-
461 sis could be advantageous in assessing the composition of the resultant model. Following

462 a detailed exploration of the biological information, addition of necessary engineering
463 assumptions and abstractions, and implementation of the computational tool, the Leish-
464 mania Petri net model comprised 174 transitions between places, with each transition
465 designed to capture a particular biological pathway. Although the authors were able to
466 show that the model could recapitulate the progression of the disease, in comparison
467 to a laboratory experimental model, and predict cellular composition within granulomas
468 [3], no analysis has been previously undertaken as to the necessity of each of the 174
469 transitions in the model. Such an analysis has the potential to infer further information
470 regarding the key biological pathways involved in disease progression and immune system
471 regulation. From an engineering perspective, the most computationally intensive process
472 in running a Petri net model is initialising each of the transitions: if a number of these
473 transitions were found to be unnecessary, there is thus potential for a large increase in
474 simulation performance.

475 To examine the impact of each of the 174 transitions, we modified the Petri net model
476 such that the simulation recorded the number of times each transition fired. As the
477 firing of a transition is potentially dependent on the initial conditions and parameter
478 values, we ran the model under a number of initial conditions, over the parameter ranges
479 originally explored by [3]. To ensure adequate coverage of the parameter space, we utilised
480 the ASPASIA sensitivity analysis toolkit [16] to generate 600 sets of parameter value
481 combinations using latin-hypercube sampling [32, 35]. By executing the Petri-Net model
482 under each of the 600 conditions, we were able to determine the number of times each
483 transition fired across the parameter space. Where a transition was found not to fire for
484 any set of initial conditions, one could question the necessity of including this pathway in
485 the model.

486 This analysis identified 47 of the 174 transitions between Petri net places that were
487 never fired (28%), suggesting a number of the transitions could potentially be removed.
488 Although this would reduce the computational complexity of the model, making simulated
489 analyses and experiments faster, it is important that we understand the impact this
490 change has in terms of our understanding of what the model captures. The argumentation
491 constructed in the development or analysis of a model provides a tool through which any
492 impact can be assessed. Of these 47 transitions that are related to T-Cells, NKT, and NK
493 cells, the majority of non-firing transitions are found to control the silencing of cells due

494 to a lack of a certain cytokine and reprise of cytokine production due to an increase in
495 environmental cytokine levels. This would suggest that the simulation is never reaching
496 thresholds where these cells are transitioning states. Knowing this, it becomes possible to
497 read through the argumentation to determine if this cell behaviour could emerge from the
498 manner in which the model has been constructed, or whether this is an error. This result
499 could also assist conversations with collaborating biologists, and provide insight into the
500 composition of the granuloma environment.

501 DISCUSSION

502 Technological advancements and a focus on interdisciplinarity has resulted in an increased
503 prevalence of laboratory studies being paired with computational modelling research,
504 motivated by the potential to reduce animal experimentation, reduce costs, and perform
505 experimentation that is not possible in the laboratory or informs future clinical studies.
506 However, for computational modelling studies to achieve that potential, it is critical that
507 the relationship between the model and the biological system being captured is fully
508 understood. Any researcher would need to have a high level of confidence in a model-
509 derived prediction before seeking to invest time, expertise, and financial resources into
510 investigating that prediction further in the real system.

511 The notion of increasing confidence in the application of computational models in
512 biological research is not new however, yet has tended to focus on the end result: the
513 implementation [27]. Such focus has led the field to suggest open-source code [26], that is
514 potentially checked by third-parties [24], and included alongside publications describing
515 that model [12]. However, the issue of confidence in a model must go further than that:
516 the code may well be adequate to do the job it has been designed to do, this does not
517 imply that the biological system has been captured appropriately [5].

518 In this tutorial we have detailed a process through which the rationale underlying
519 the design, implementation, and analysis of a model of a biological system is generated.
520 We see this process being applied either within a process of model construction or as
521 a tool through which an assessment of a previously developed model can be performed.
522 This process begins by examining the purpose of the study: what it is that the model
523 will be used for. This establishes the scope prior to any experimental work, to ensure

524 the tool is not being used to generate predictions for which it has not been designed.
525 This purpose is then a key consideration in an examination of each component phase
526 of model development: assessing the biological data; making necessary assumptions in
527 place of a lack of information; choosing the correct modelling paradigm; introducing
528 necessary modelling assumptions; engineering the computational model, and performing
529 experimentation using the tool. Any omissions or ambiguities inherent in any of these
530 phases could impact the potential to relate a model prediction to the real-world: for this
531 to be detected, all design decisions must be transparent.

532 Adverse outcome pathway (AOP) tools have found application in toxicology and in
533 studies of human risk assessment, providing a means to specify how interactions at the
534 molecular, cellular and organ level can be linked to an adverse outcome [38]. Presented
535 as a flow diagram, AOPs can show the strength of evidence supporting the events in
536 the outcome pathway, yet have come under criticism for splitting the representation of
537 the process from the evidence, providing a simplistic representation of the toxicological
538 process [38]. More generally, yet applicable to QSP-related models, the ODD (Overview,
539 Design concepts, Details) protocol does permit the specification of the purpose behind
540 the creation of a model, the inclusion of biological components, and modules describing
541 the implementation of biological behaviour, alongside relevant assumptions [21]. The
542 focus of ODD is scientific repeatability, rather than fitness for purpose as specified in this
543 tutorial, and lacks the recording of model experimentation and statistical analyses, and
544 motivation for performing those experiments [21]. In producing this tutorial we are not
545 hoping to replace either technique: argumentation could be used alongside either, but
546 we do contend that neither method provides the complete set of information required to
547 convince researchers that a model is appropriately constructed and analysed to meet its
548 intended purpose.

549 We believe that arguing over the rationale for each of the model development phases
550 identified in Figure 3 can provide a transparent evidence base upon which the contribution
551 of a computational model can be assessed. Alongside a description of the process involved
552 in examining the rationale at each phase, we have shown an example application of the
553 process in examining the rationale underlying the development of a model of Leishmani-
554 asis: developed to further understand this neglected tropical disease to generate insights
555 that could inform future therapeutic studies [3]. In addition to exposing the rationale

556 behind this model, we then described how this argument could potentially be used to de-
557 termine the links between this model and human visceral leishmaniasis (HVL), and how
558 the argument could be useful in examining the composition of the model with respect to
559 computational complexity. The approach offers more than a process to be employed in
560 model development or assessment, and is advantageous in redefining the purpose of, or
561 refining the composition of, models developed for QSP studies. Where a computational
562 model is closely tied to a mouse study, structured argumentation using the approach de-
563 tailed in this tutorial has the potential to provide a robust way of understanding how the
564 model could be repurposed for human studies that predate or inform clinical trials.

565 Acknowledgements

566 This work is part funded by the Crack-IT programme, grant number NC/C013117/1 and
567 NC/C013205/1. PMK would also like to acknowledge support from MRC Programme
568 Grant G1000230. JT would like to acknowledge support from The Royal Academy of
569 Engineering and The Royal Society.

570 References

- 571 [1] Gsn working group. Accessed: 2016-09-18.
- 572 [2] Combining software evidence arguments and assurance. In *Proceedings of ICSE-*
573 *2005: Workshop on Realising Evidence Based Software Engineering* (2005).
- 574 [3] ALBERGANTE, L., TIMMIS, J., BEATTIE, L., AND KAYE, P. M. A Petri net model
575 of granulomatous inflammation: implications for IL-10 mediated control of Leishma-
576 nia donovani infection. *PLoS computational biology* 9, 11 (jan 2013), e1003334.
- 577 [4] ALDEN, K., ANDREWS, P. S., POLACK, F. A. C., VEIGA-FERNANDES, H.,
578 TIMMIS, J., AND COLES, M. C. Using Argument Notation to Engineer Biological
579 Simulations with Increased Confidence. *Journal of the Royal Society Interface* 12,
580 105 (2015).
- 581 [5] ALDEN, K., AND READ, M. Computing: Scientific software needs quality control.
582 *Nature* 502 (2013), 448.

- 583 [6] ALDEN, K., READ, M., TIMMIS, J., ANDREWS, P. S., VEIGA-FERNANDES, H.,
584 AND COLES, M. C. Spartan: A Comprehensive Tool for Understanding Uncertainty
585 in Simulations of Biological Systems. *PLoS computational biology* 9, 2 (2013).
- 586 [7] ALDEN, K., TIMMIS, J., ANDREWS, P. S., VEIGA-FERNANDES, H., AND COLES,
587 M. C. Pairing experimentation and computational modelling to understand the
588 role of tissue inducer cells in the development of lymphoid organs. *Frontiers in*
589 *Immunology* 3 (2012), 1–20.
- 590 [8] ALVAR, J., VLEZ, I. D., BERN, C., HERRERO, M., DESJEUX, P., CANO, J.,
591 JANNIN, J., BOER, M. D., AND THE WHO LEISHMANIASIS CONTROL TEAM.
592 Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* 7, 5 (05
593 2012), 1–12.
- 594 [9] AN, G. Agent-Based Computer Simulation and SIRS: building a bridge between
595 basic science and clinical trials. *Shock* 16, 4 (2001), 266–273.
- 596 [10] AN, G. Concepts for developing a collaborative in silico model of the acute inflam-
597 matory response using agent-based modeling. *Journal of critical care* 21, 1 (mar
598 2006), 105–10; discussion 110–1.
- 599 [11] ANDREWS, P. S., POLACK, F., SAMPSON, A. T., TIMMIS, J., SCOTT, L., AND
600 COLES, M. Simulating biology : towards understanding what the simulation shows.
601 In *Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation*
602 (2008), pp. 93–123.
- 603 [12] BARNES, N. Publish your computer code: it is good enough. *Nature* 467 (2010),
604 753.
- 605 [13] CHINNECK, P., PUMFREY, D., AND McDERMID, J. The heat/act preliminary
606 safety case: A case study in the use of goal structuring notation. In *Proceedings of*
607 *the 9th Australian Workshop on Safety Critical Systems and Software - Volume 47*
608 (Darlinghurst, Australia, Australia, 2004), SCS '04, Australian Computer Society,
609 Inc., pp. 33–41.

- 610 [14] CILFONE, N. A., KIRSCHNER, D. E., AND LINDERMAN, J. J. Strategies for effi-
611 cient numerical implementation of hybrid multi-scale agent-based models to describe
612 biological systems. *Cellular and Molecular Bioengineering* 8, 1 (2015), 119–136.
- 613 [15] DENNEY, E., PAI, G., AND HABLI, I. Towards measurement of confidence in
614 safety cases. In *2011 International Symposium on Empirical Software Engineering*
615 and Measurement (Sept 2011), pp. 380–383.
- 616 [16] DYSON, S., ALDEN, K., CUCURULL-SANCHEZ, L., LARMINE, C., COLES, M.,
617 KULLBERG, M., AND TIMMIS, J. Aspasia: A toolkit for evaluating the effects of
618 biological interventions on sbml model behaviour. *In Review*.
- 619 [17] EXECUTIVE, H., AND SAFETY. *Reducing risks, protecting people*. Crown, 2001.
- 620 [18] FENN, J., HAWKINS, R., WILLIAMS, P., AND KELLY, T. *The Safety of Systems:*
621 *Proceedings of the Fifteenth Safety-critical Systems Symposium, Bristol, UK, 13–15*
622 *February 2007*. Springer London, London, 2007, ch. Safety Case Composition Using
623 Contracts - Refinements based on Feedback from an Industrial Case Study, pp. 133–
624 146.
- 625 [19] GAZZINELLI, R., OSWALD, I., JAMES, S., AND SHER, A. Il-10 inhibits parasite
626 killing and nitrogen oxide production by ifn-gamma-activated macrophages. *Journal*
627 *of Immunology*, 148 (1992), 1792–1796.
- 628 [20] GHETIU, T., POLACK, F. A. C., AND BOWN, J. Argument-driven Validation
629 of Computer Simulations A Necessity Rather Than an Option. In *VALID 2010:*
630 *The Second International Conference on Advances in System Testing and Validation*
631 *Lifecycle* (2010), pp. 1–4.
- 632 [21] GRIMM, V., BERGER, U., BASTIANSEN, F., ELIASSEN, S., GINOT, V., GISKE,
633 J., GOSS-CUSTARD, J., GRAND, T., HEINZ, S. K., HUSE, G., HUTH, A.,
634 JEPSEN, J. U., JØRGENSEN, C., MOOIJ, W. M., MÜLLER, B., PEER, G., PIOU,
635 C., RAILSBACK, S. F., ROBBINS, A. M., ROBBINS, M. M., ROSSMANITH, E.,
636 RÜGER, N., STRAND, E., SOUSSI, S., STILLMAN, R. A., VABØ, R., VISSER, U.,
637 AND DEANGELIS, D. L. A standard protocol for describing individual-based and
638 agent-based models. *Ecological Modelling* 198, 1-2 (sep 2006), 115–126.

- 639 [22] GUO, Z., AND TAY, J. C. A Comparative Study on Modeling Strategies for Im-
640 mune System Dynamics Under HIV-1 Infection. In *Artificial Immune Systems, 4th*
641 *International Conference, ICARIS 2005, LNCS 3627* (2005), Springer, pp. 220 – 233.
- 642 [23] HAMILL, P. *Unit Test Frameworks: Tools for High-Quality Software Development.*
643 ” O'Reilly Media, Inc.”, 2004.
- 644 [24] HAYDEN, E. C. Mozilla plan seeks to debug scientific code. *Nature* 501, 472 (2013).
- 645 [25] I, K., W, R., AND F, S. *Modeling in Systems Biology: The Petri Net Approach.*
646 Springer, 2011.
- 647 [26] INCE, D. C., HATTON, L., AND GRAHAM-CUMMING, J. The case for open com-
648 puter programs. *Nature* 482, 7386 (feb 2012), 485–8.
- 649 [27] JOPPA, L. N., MCINERNY, G., HARPER, R., SALIDO, L., TAKEDA, K., O'HARA,
650 K., GAVAGHAN, D., AND EMMOTT, S. Troubling Trends In Scientific Software Use.
651 *Science* 340 (2013), 814–815.
- 652 [28] KANE, M., AND MOSSER, D. The role of il-10 in promoting disease progression in
653 leishmaniasis. *Journal of Immunology* 166 (2001), 1141–1147.
- 654 [29] KAYE, P. M., SVENSSON, M., ATO, M., MAROOF, A., POLLEY, R., STAGER,
655 S., ZUBAIRI, S., AND ENGWERDA, C. R. The immunopathology of experimental
656 visceral leishmaniasis. *Immunological Reviews* 201, 1 (2004), 239–253.
- 657 [30] KELLY, T. P. *Arguing Safety A Systematic Approach to Managing Safety Cases.*
658 PhD thesis, University of York, YCST 99/05, 1999.
- 659 [31] MCCALL, L.-I., AND MATLASZEWSKI, G. Involvement of the leishmania donovani
660 virulence factor {A2} in protection against heat and oxidative stress. *Experimental*
661 *Parasitology* 132, 2 (2012), 109 – 115.
- 662 [32] MCKAY, M. D., BECKMAN, R. J., AND CONOVER, W. J. A comparison of
663 three methods for selecting values of input variables in the analysis of output from a
664 computer code. *Techometrics* 21 (1979), 239–245.
- 665 [33] MURRAY, H. W. Tissue granuloma structure-function in experimental visceral leish-
666 maniasis. *International Journal of Experimental Pathology* 82, 5 (2001), 249–267.

- 667 [34] POLACK, F. A. C., DROOP, A., GARNETT, P., GHETIU, T., AND STEPNEY, S.
668 Simulation validation : exploring the suitability of a simulation of cell division and
669 differentiation in the prostate. In *CoSMoS 2011: Proceedings of the 2011 Workshop*
670 on *Complex Systems Modelling and Simulation* (2011).
- 671 [35] READ, M., ANDREWS, P. S., TIMMIS, J., AND KUMAR, V. Techniques for
672 Grounding Agent-Based Simulations in the Real Domain : a case study in Experi-
673 mental Autoimmune Encephalomyelitis. *Mathematical and Computer Modelling of*
674 *Dynamical Systems* 18, 1 (2012), 67–86.
- 675 [36] ROSENBERG, L. H., THEODORE, P., HAMMER, F., AND HUFFMAN, L. L. Re-
676 quirements, testing and metrics. In *In 15th Annual Pacific Northwest Software Qual-*
677 *ity Conference* (1998).
- 678 [37] VERMA, S., KUMAR, R., KATARA, G. K., SINGH, L. C., NEGI, N. S., RAMESH,
679 V., AND SALOTRA, P. Quantification of parasite load in clinical samples of leish-
680 maniasis patients: Il-10 level correlates with parasite load in visceral leishmaniasis.
681 *PLoS ONE* 5, 4 (04 2010), 1–8.
- 682 [38] VINKEN, M. The adverse outcome pathway concept: a pragmatic tool in toxicology.
683 *Toxicology* 312 (oct 2013), 158–65.

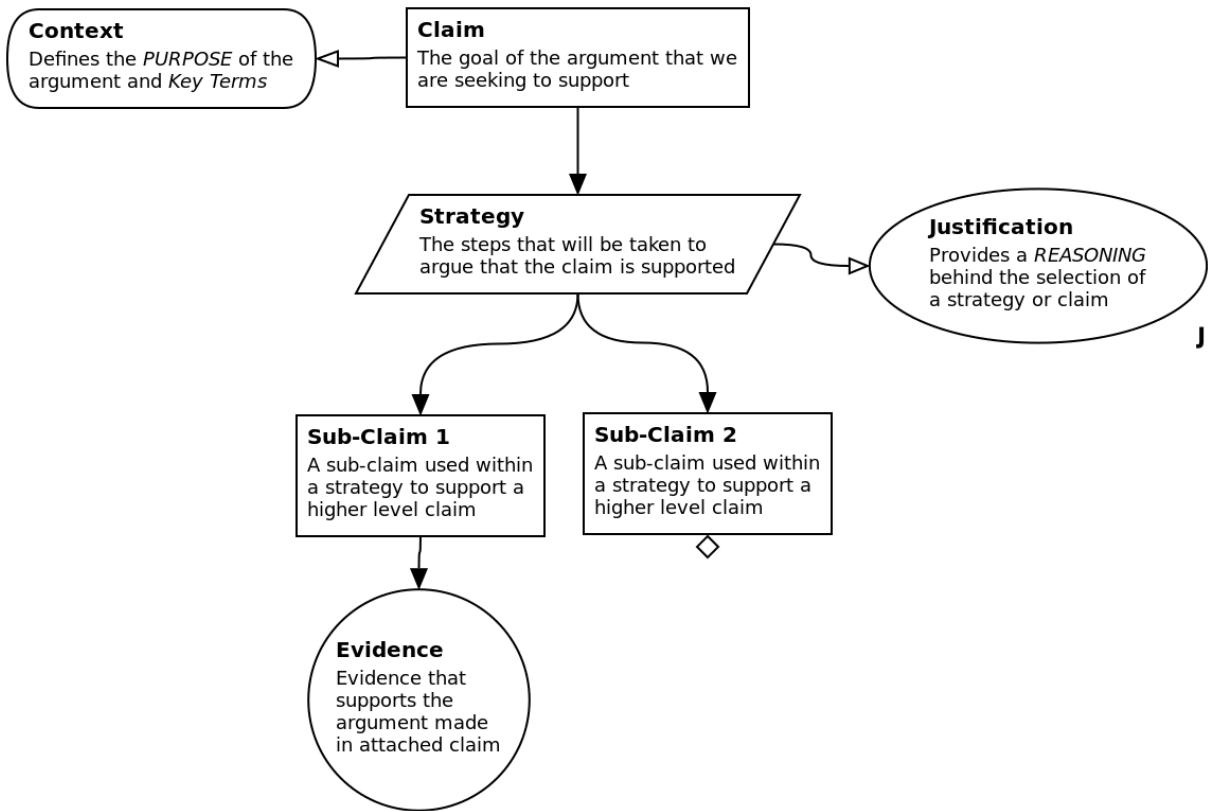


Figure 1: Description of the notation used in the creation of an argument.

To show how these components are linked together, we present the description of each component within the format of an argument structure.

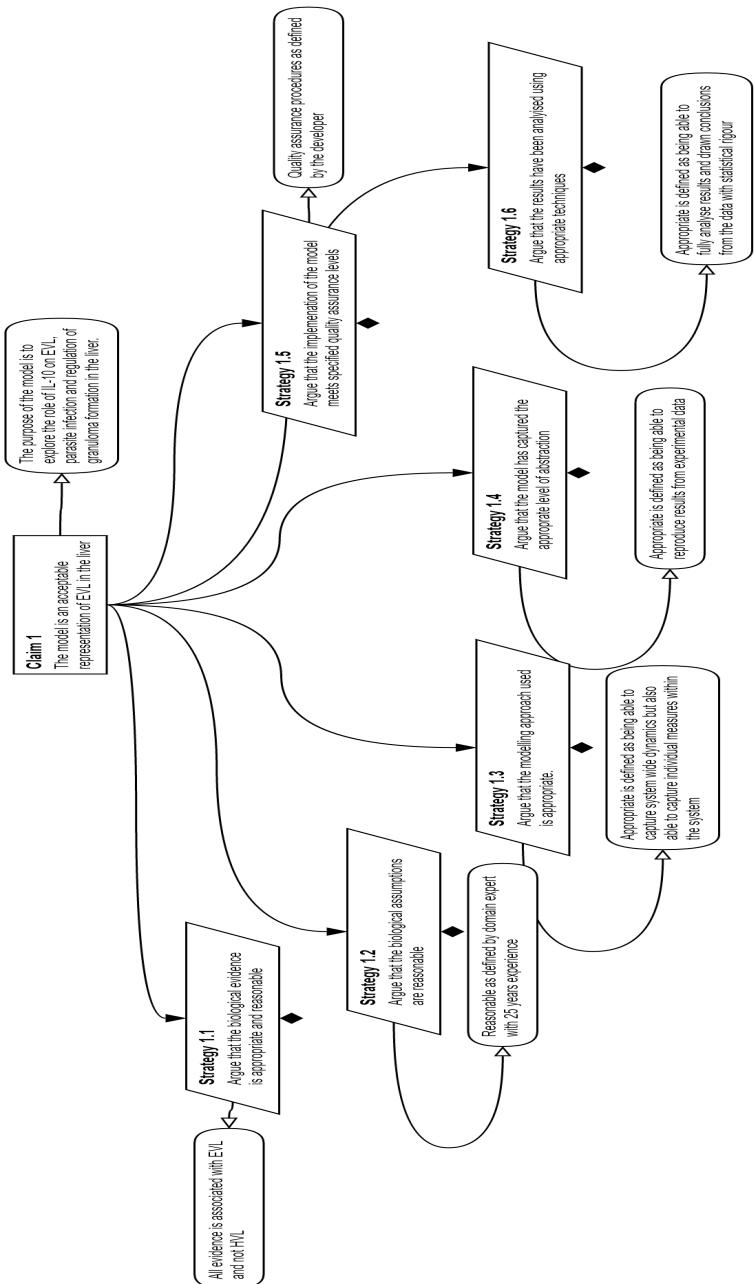


Figure 2: Top level argument in the process of arguing that the leishmaniasis simulation is fit for purpose. Black diamonds indicate the strategies in this figure are expanded upon below.

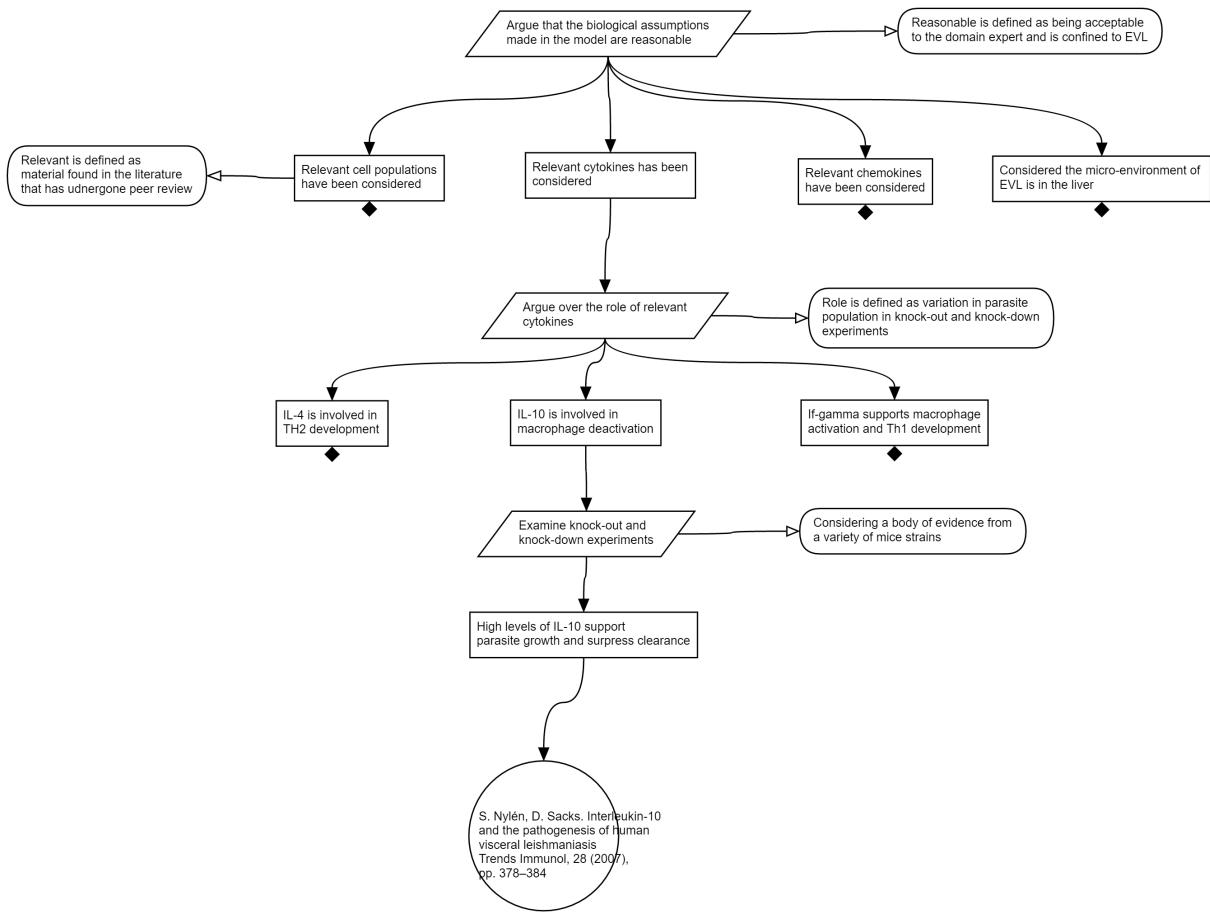


Figure 3: Arguing appropriateness of evidence used as a basis for the Leishmaniasis Simulation in [3].

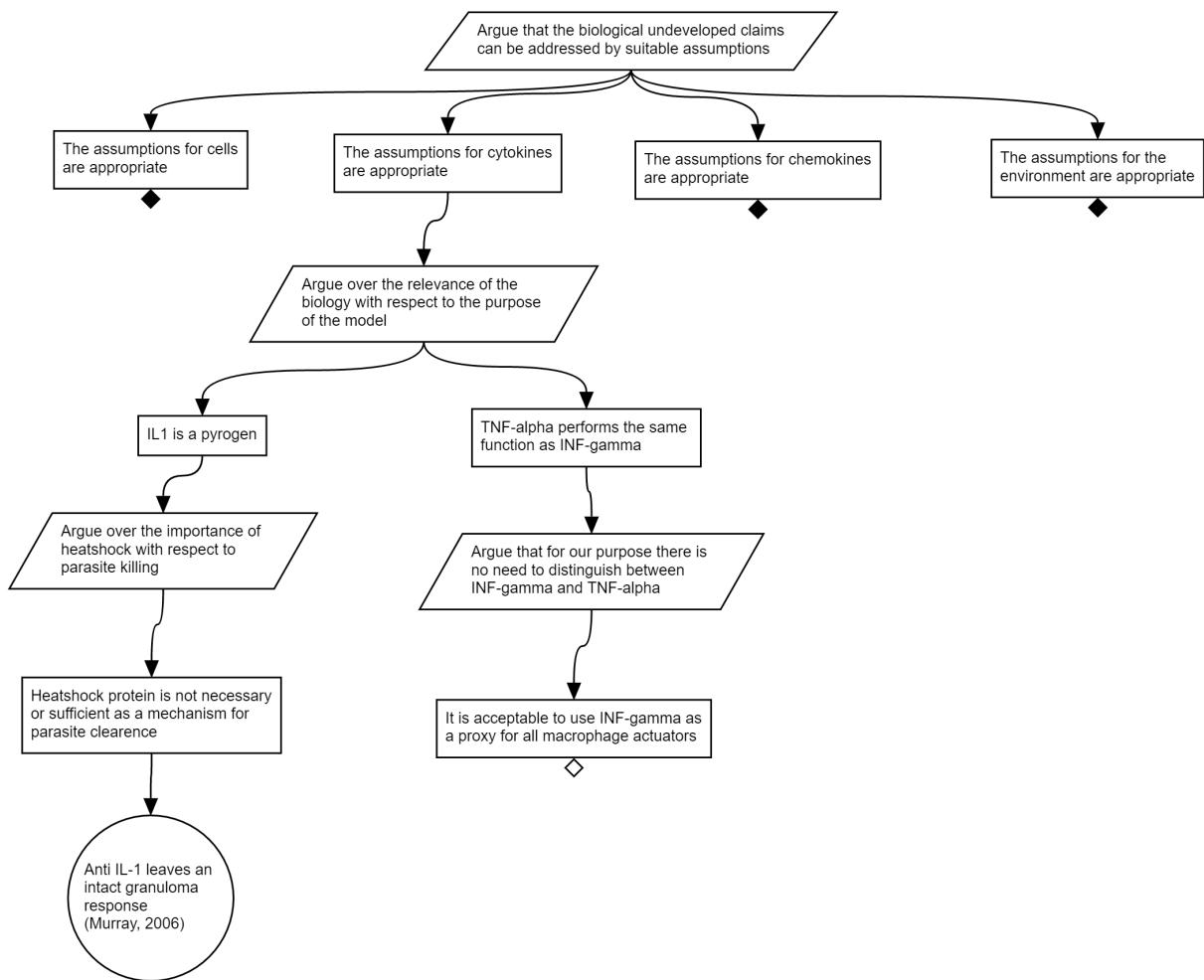


Figure 4: Argument that the biological abstractions introduced in the model are suitable. In this case, the approach is exemplified by focusing on abstractions of cell type to be included in the model.

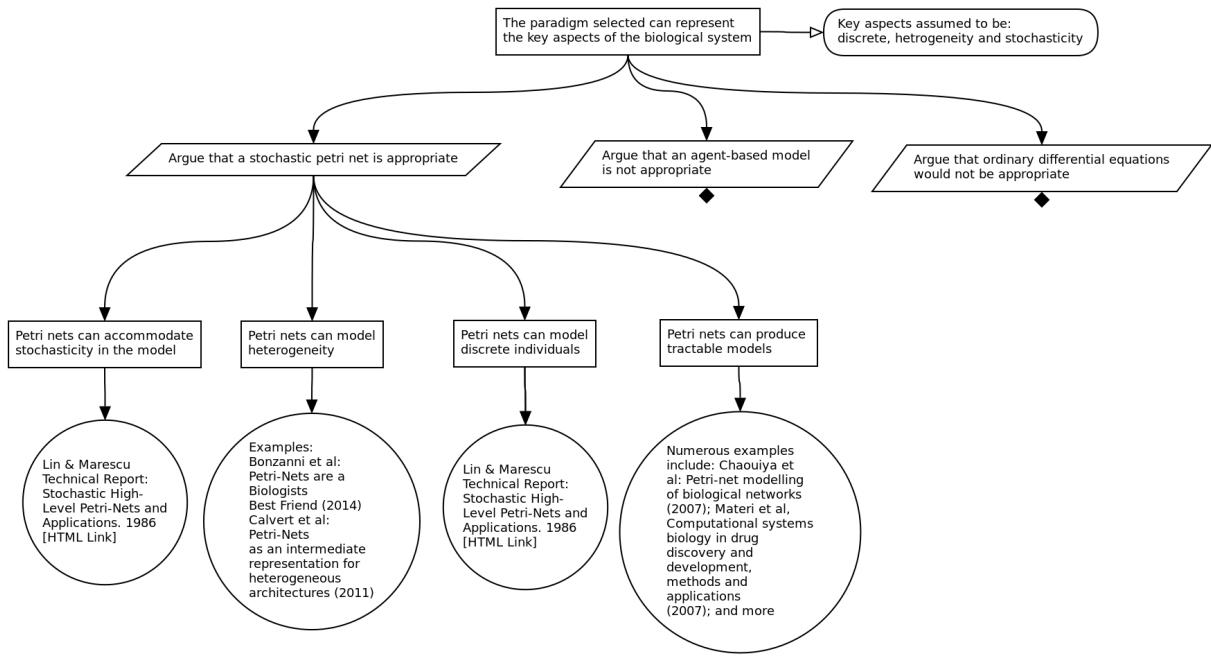


Figure 5: Argument that the adopted modelling approach is adequate given the research context. In this case, the approach is exemplified by focusing on the choice of modelling paradigm: Petri Net, Agent-Based model, or ODEs.

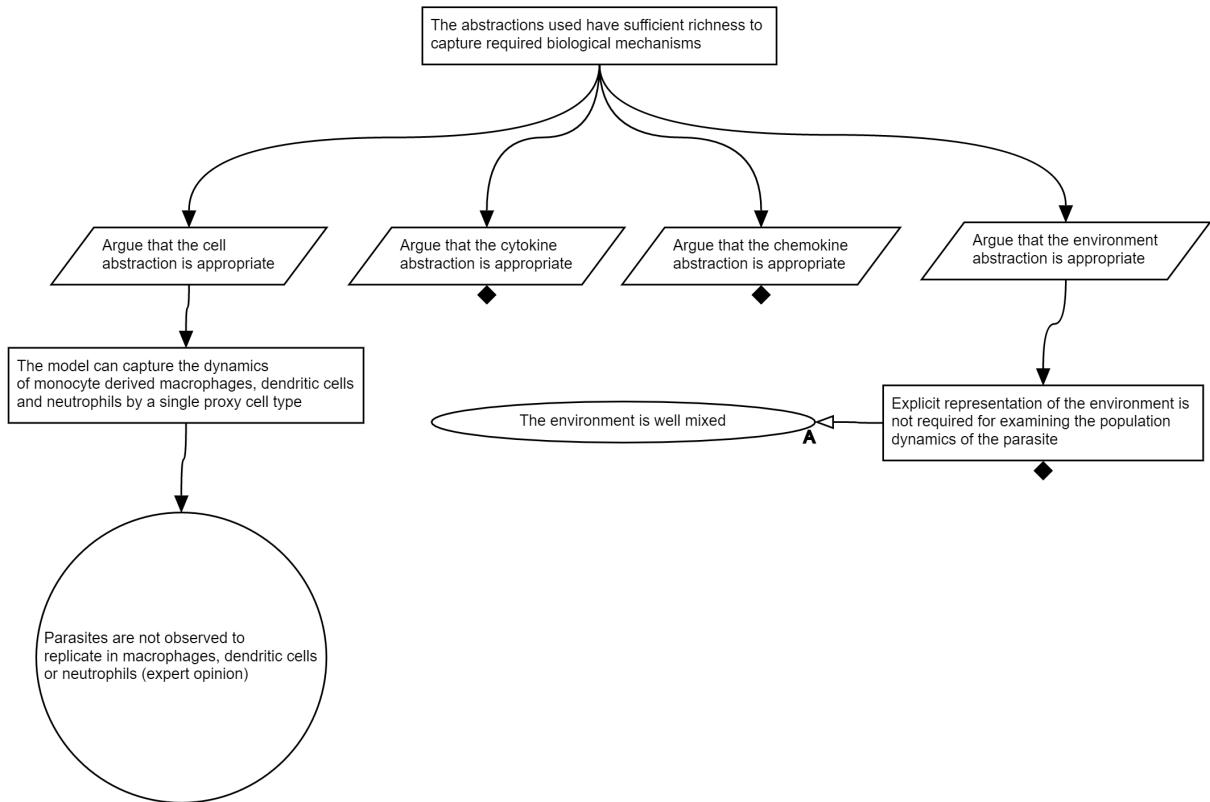


Figure 6: Subset of the argument that supports the rationale for abstracting modelling abstractions.

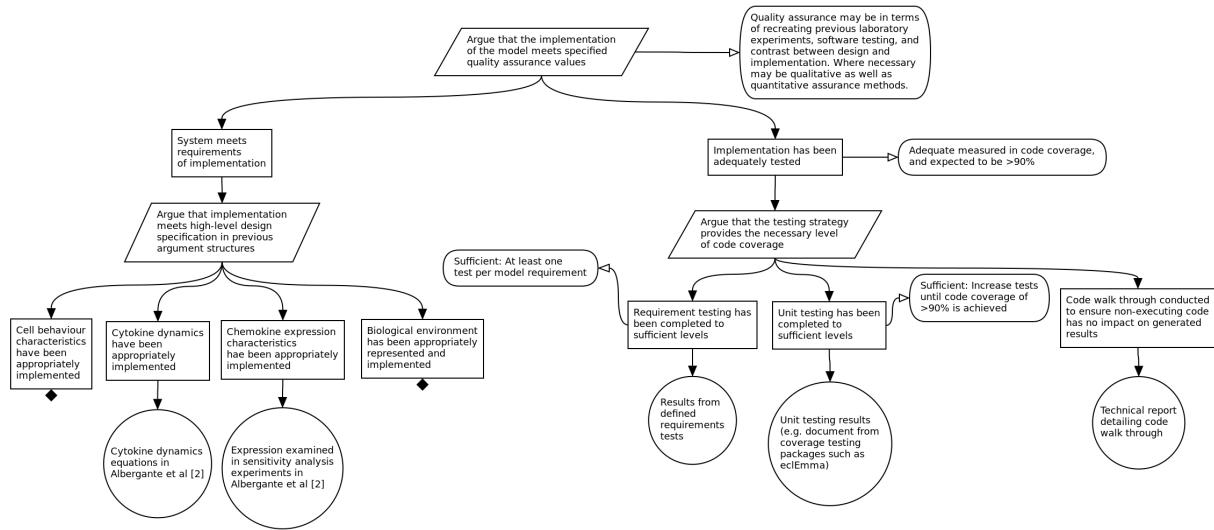


Figure 7: Subset of the argumentation structure used to argue that the implementation of the model is adequate for meeting the purpose specified in Figure 2.

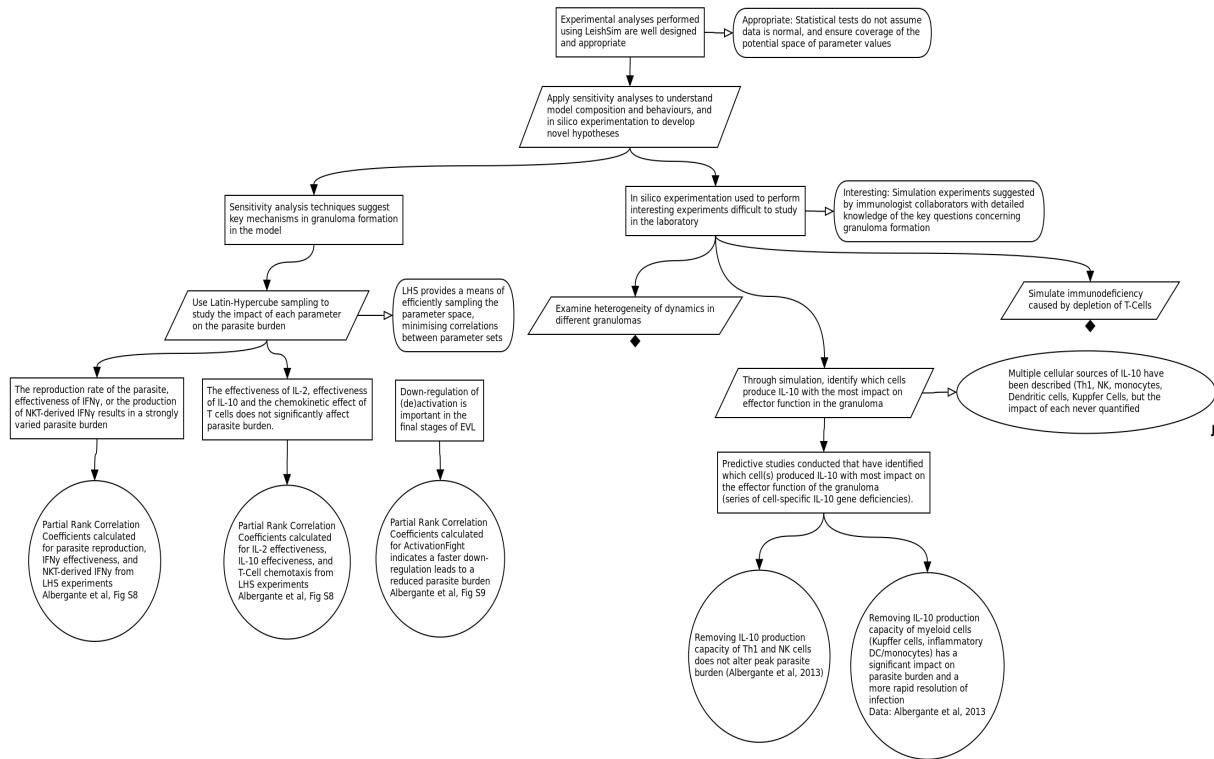


Figure 8: Subset of the argumentation structure for the design of the experimental analysis.

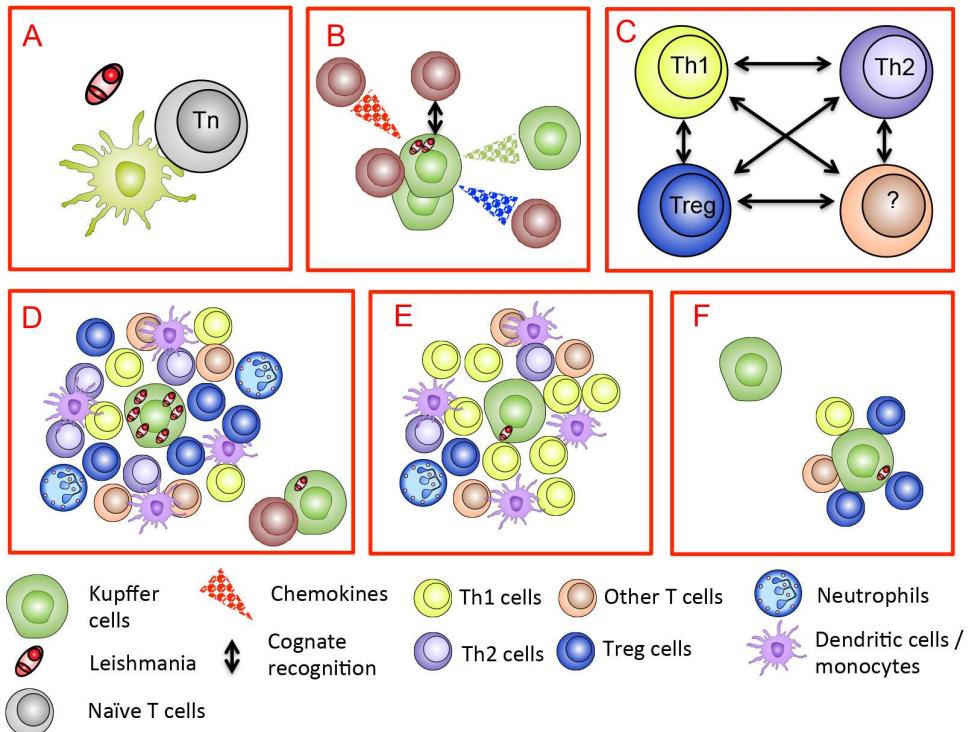


Figure S1: Summary of immune response associated with granuloma formation in leishmaniasis. A. Within hours of experimental infection with *Leishmania donovani*, dendritic cells present parasite antigens to naïve T lymphocytes in lymphoid tissues to initiate an adaptive immune response. B. Simultaneously, parasites in the liver infect resident liver macrophages (Kupffer cells), stimulating the production of chemokines that attracts innate lymphoid cells (of which NKT cells are best characterized). NKT cells engage with infected Kupffer cells via cognate receptor-ligand interactions, amplifying the chemokine response to attract additional Kupffer cells, NKT cells and eventually other cell types (see D, below). C. Over the first few days of infection, T cells differentiate into a variety of subsets (Th1, Th2, Treg), producing cytokines that may cross-inhibit or cross-stimulate T cell differentiation. These cytokines also promote (e.g. IFN) or inhibit (e.g. IL-10) the ability of macrophages to kill *Leishmania*. D. The relative balance of different T cell subsets, together with monocytes, dendritic cells, and occasionally neutrophils that are attracted to the expanding granuloma determines parasite burden. Notably, granuloma development is asynchronous (lower right). E. Reduction in parasite burden is achieved when Th1-type immune responses become dominant. F. Resolution of infection is accompanied by granuloma involution (loss of cellularity) and a restoration of homeostasis. Experimental and modeling data suggest, however, that some residual parasites survive in some granulomas due to regulatory mechanisms.

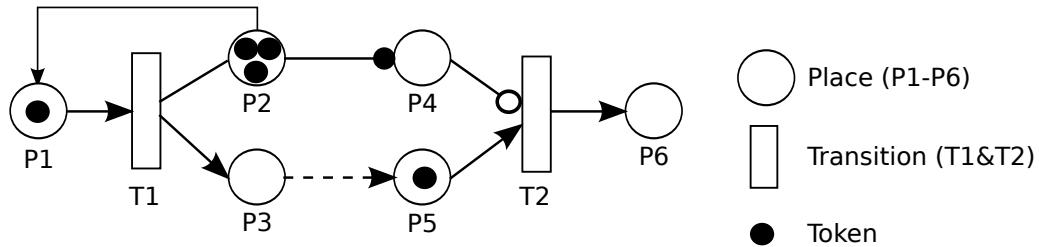
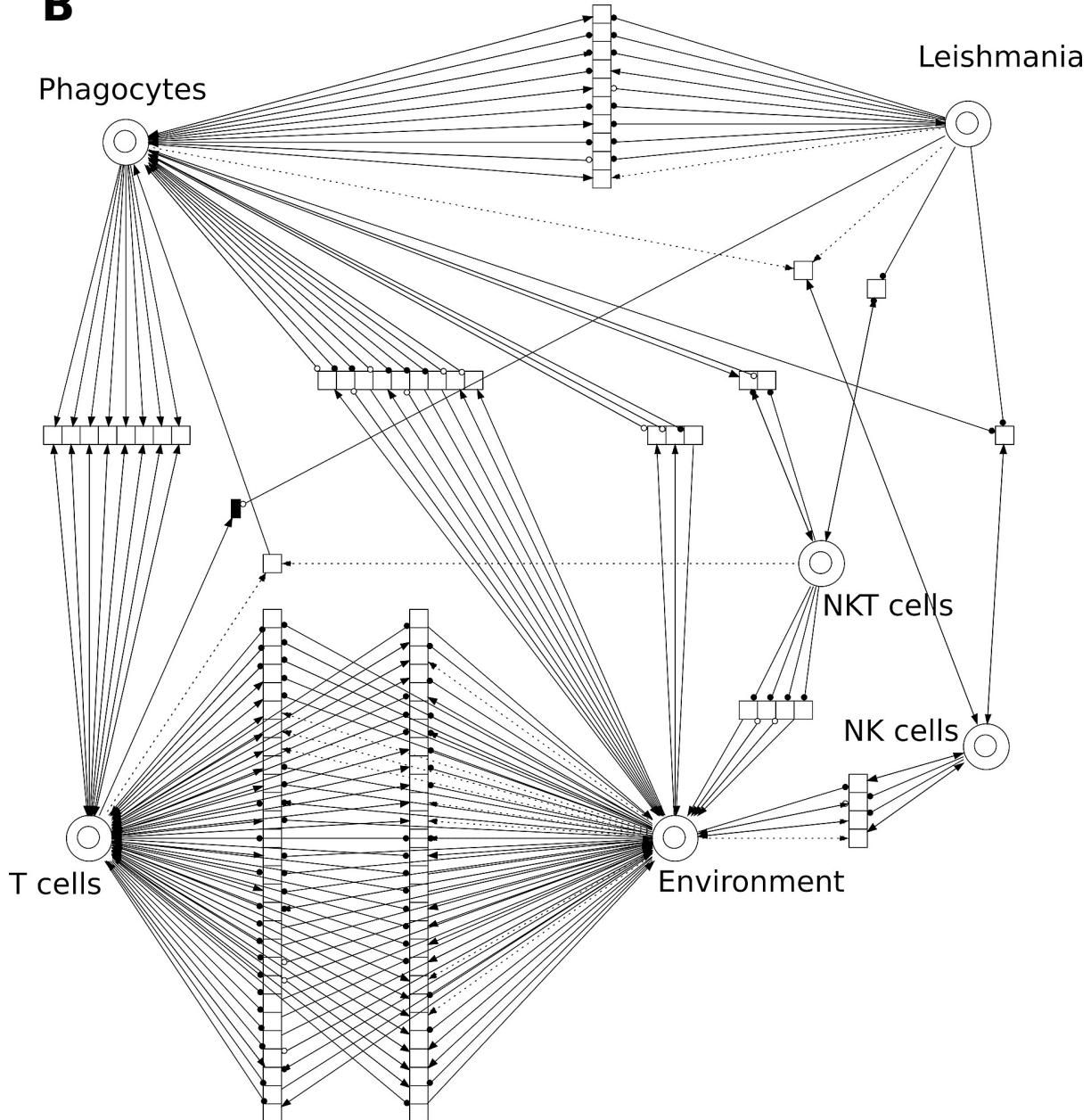
A**B**

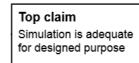
Figure S2: Petri-Net modeling approach used to develop the case study model of granuloma formation in Leishmaniasis. A: Schematic of Petri net Places (P1,2,3,4), tokens (black circles in places) and transitions (T1,2). Continuous line, Standard arrowhead: takes tokens from the input places and moves tokens to the output place. Dotted Line, Standard arrowhead: The number of tokens of a place is used in the evaluation of the rate of a transition. Continuous line, Full circle: Target transition only performed if the appropriate number of tokens is present in input. Continuous line, Empty circle: Disables the target transition if the appropriate number of tokens is present in the input place. B: High level Petri-Net model of granuloma formation, reproduced from [3].

Semantics Used in Artoo Argumentation Tool			
Notation	Definition	Description	Connected To
	Claim	A Claim is an identified fitness-for-purpose requirement that the argument is seeking to substantiate, if possible. As an argument around a claim is constructed, that argument can be broken down into subsets of claims that, if substantiated, support the substantiation of the higher level claim	→ → → Context → Assumption → Strategy → Evidence
	Strategy	A Strategy node should state the specific actions that have been taken to substantiate the claim to which this node is attached. This strategy may consist of breaking the claim down into a subset of claims which are then argued in turn.	→ → Assumption → Justification → Sub-Claim → Unsubstantiated → Claim → Claim Continued
	Context	A Context node should be used to provide contextual information concerning information in a node to which it is attached. This information may be definition of particular words or phrases (such as adequate) or the level at which the attached claim is deemed to have been substantiated.	
	Assumption	An Assumption node provides a means of specifying any information that is assumed to be true when arguing over a claim or designing a strategy to examine a particular claim. Explicitly stating the inherent assumptions eases the process by which others can assess the extent to which the argument over a particular claim holds.	
	Justification	A Justification node should contain the reasoning for the application of a particular strategy in order to substantiate a claim. Justifying the approach used explicitly can reveal the extent to which alternative strategies have been considered, and why this strategy was selected over those alternatives.	
	Unsubstantiated Claim	Unlike the application of structured notation in formal safety-case arguments, a claim can be shown to be unsubstantiated in the approach described in this paper. Biological systems are not fully understood, and it may not be possible to generate evidence to substantiate the complete set of fitness for purpose requirements. Where this is the case, it is critical that the lack of evidence is explicitly stated in the argument, and the limitations of the model are shown. In our approach, a lack of substantiating evidence is shown by attaching a white diamond to that claim	
	Claim Continued	As the argument becomes more complex, it may become difficult to follow. As such we have introduced a black diamond notation, representing the continuation of the argument surrounding this claim on a different diagram	
	Evidence	Node containing the Evidence that is used to substantiate an attached claim. In Artoo, it is possible to hyperlink to this evidence, which could include publications, experimental results, statistical analyses, etc.	

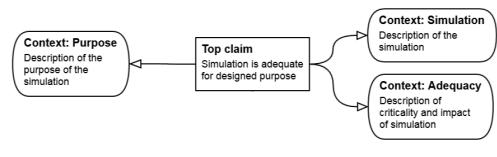
Figure S3: Semantics of diagram language used in Artoo

Box 1: Developing a Claim

A Identify a claim that the argument seeks to support. This is represented within a rectangle. The objective is to detail how this claim can be supported with available evidence, if possible.

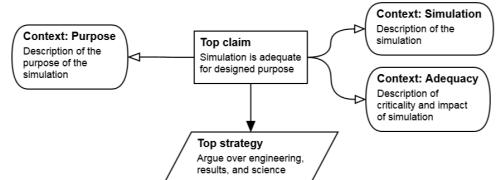


B Each claim is usually examined within a given context. This may involve defining the meaning of key terms stated in the claim. For example, if the purpose of the model was to adequately capture a biological system, adequate must be defined. Any context definitions are given in rounded rectangles

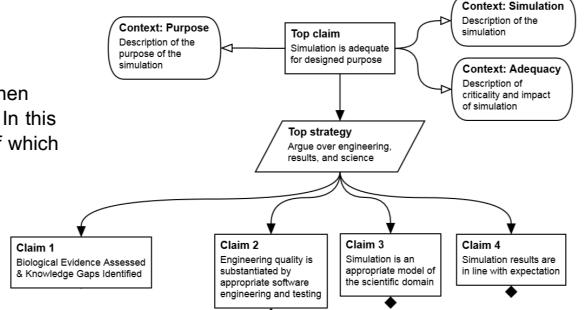


C Each claim is accompanied by one or more strategies that will be used to determine if that claim can be supported. This could, as examples, be a particular experimental strategy or systematic literature review. A strategy is always stated in a parallelogram.

A claim or strategy can also be accompanied by a justification or assumption node to provide more detail on the choice of the claim being made or the strategy that was followed. Semantics in Supplementary Figure 1



D Strategies, unless leading to evidence (see part E), are then broken down into sub-claims, and the process repeated. In this case, the strategy is divided into four sub-claims, each of which examines a key part of the model development process.



E If evidence can be provided that supports a claim, this is stated in an evidence node: a circle. In Artoo, electronic links to this evidence can be provided.

Diamonds on the diagram indicate either:

(i) a claim cannot be supported. If no evidence can be provided, a white diamond can be used to show this is the case.

(ii) the argument is detailed on another diagram. A black diamond is used to show the claim is fully described elsewhere.

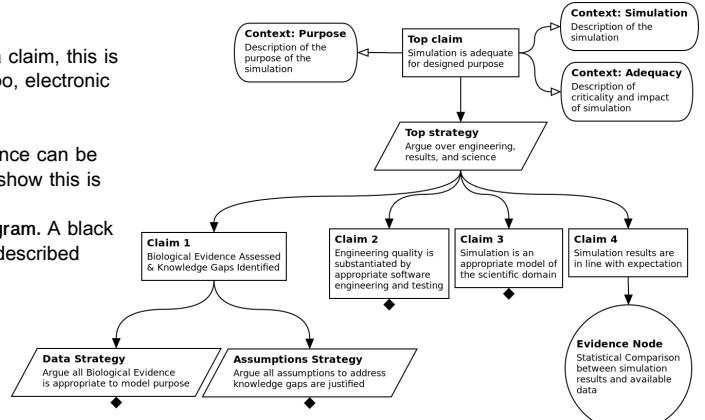


Figure S4: Process of developing a specific claim, using the diagrammatic notation used in Artoo.

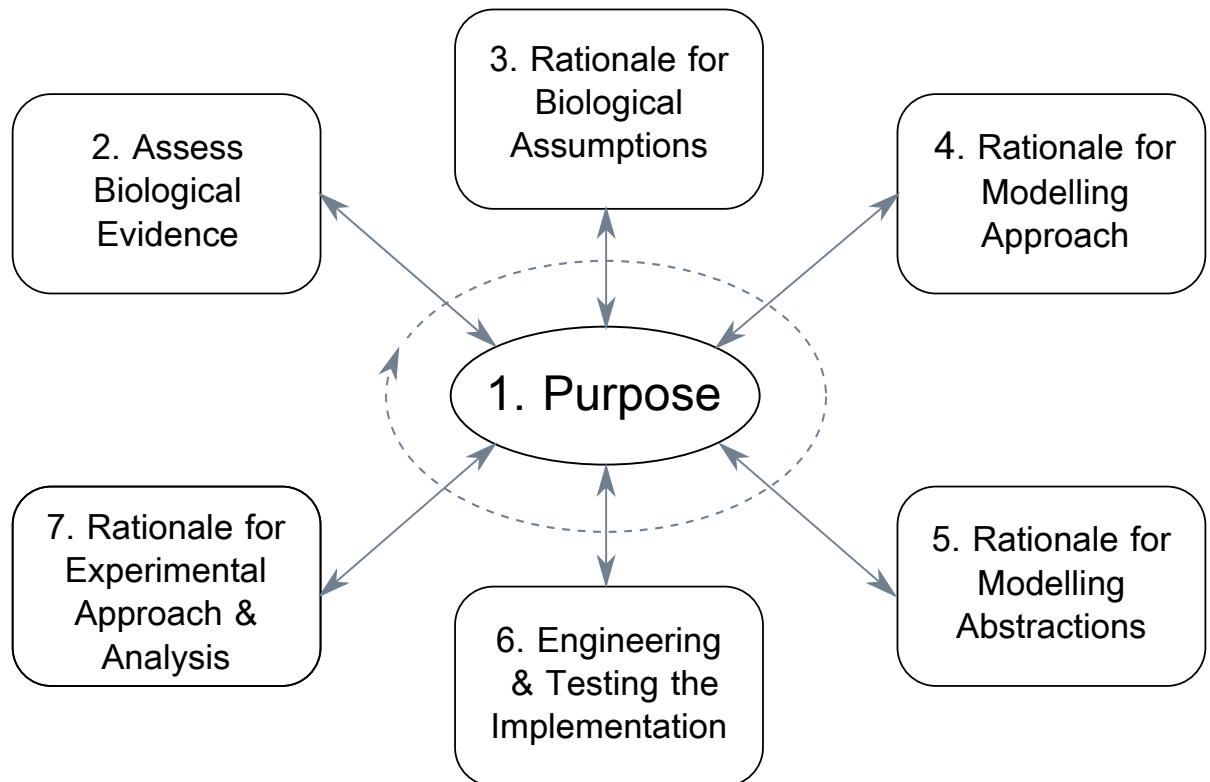


Figure S5: Process through which assessing the rationale for model design, implementation, and analysis should be conducted. Each stage of the process is grounded in the purpose for which the model was developed. Arrows linking to Purpose are bidirectional as the purpose shapes what assumptions and abstractions are appropriate, and conversely, decisions about assumptions and abstractions that are made can de facto alter the purpose for which the model is fit. Note the lack of defined end point: arguing fitness for purpose has potential to inform later iterations of model and study development.