Appendix C

GA User Questionnaire Responses

This appendix presents the individual anonymised responses received in the GA user study presented in Chapter 3.

GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - TI

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

TI : 7 years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?


*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

TI : Interest in evolution. Interest in seeing how well GAs work on such problems. General research in optimisation and landscape structures.
*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAS? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY i.e. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

TI:
Unix
gcc 2.6.3
C++ language
Tcl/Tk for scripting and interfaces
Tcl-dp for distributed purposes
Plplot for graph drawing (not much yet)
also macintosh with symantec c++ for development purposes

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

TI: This is to my mind the most important step in any algorithm, perhaps more important than the choice of algorithm.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

TI: I'm less interested in this - as I generally look at pretty precise TSP problems or whatever, and investigate landscapes and other such things genetic

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

TI: genetic operator very important and hard to pick. The rest ain't too important in my opinion
*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

TI : No opinion

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

TI :

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

TI : Extract loads of data during runs – not very efficient, but I want to know as much as possible.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

TI : Compare with other algorithms, known bounds etc. This is important to me.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

TI : none for my purposes

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

TI : too much to look at
*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

TI:

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

TI:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

TI: Need more than this; need to know the local structure of fitness changes throughout the population.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

TI:

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

TI: I'm more interested in more general questions and different algorithms, so this ain't much use to me.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

TI:

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

TI:


TI:


TI:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOC") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

TI:

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

TI:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

TI: The ratio of accepted changes to non-accepted, the correlation in fitness values as a function of distance in representation space, the frequency of local fitness peaks, ... comparisons with other algorithms.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

TI: Best off just grabbing all the data and looking at it later

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

TI: No

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

TI: No

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

TI: No

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.
*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

TI: I’d like to see any tool be applicable more broadly than just to GAs, but then I don’t care too much about chromosomes and all that.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

TI: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - T2

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

T2 : 3 years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

T2 : Research on representation and role of mutation. Self adaption and solving specific problems.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

T2 : Research

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

T2 : Smalltalk V on Dos.
Gnu C++ on unix

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

T2 : String representation is limiting. Not useful for all problems
Better representations exist.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
Only when there are conflicting criteria.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

T2:

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

T2: This is a difficult problem, as parameter settings drastically affect the efficiency of the GA.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFicultIES YOU ENCOUNTER BELOW.

T2:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

T2: Pick an instance of the problem with a known solution, so that you can verify that if can be found. Then gather statistics on solving the problems over a number of runs.

IS THERE ANY OTHER WAY?

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

T2:
*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

T2: Can do it already

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

T2:

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

T2: See above

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

T2:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

T2: Can do it already. Shows you what is happening during single runs.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

T2: Slow down the run

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED
FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

T2 : Can do it. Only useful to check if GA works correctly, and see the effect of hamming cliffs

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

T2 :

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

T2 : The showing the occurrence of mutation by itself is not very useful. Percentage of mutations that are better than its "parent" are useful to show the effectiveness of mutation over the run.

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

T2 :


T2 : Can't see any, so have not done it.


T2 :
*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

*T2: Only bother in terms of fitness. This would be seen by the average fitness being very close to the maximum fitness.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

*T2: 

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

*T2: 

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

*T2: Can already do it.

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

*T2: Can already do it

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:
T2: Could do it, but can’t see any reason to do it.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

T2: Can not see any good reason to do it.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

T2:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

T2: Any such GA package needs to able to show visualisation of individual runs, and gather statistical info on batches of runs. Also the representation of genes and selection of which reproduction operators must be easily changed.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

T2: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - T3

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

T3 : About 4 years.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

T3 : Various timetabling and scheduling problems, real and contrived. Facility layout problems, set-covering problems, pipe-routing, and various miscellany.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

T3 : Primarily because evolutionary algorithms are my principal research interest. For practical problems, they promise flexibility and fast prototyping, though not necessarily best results of course; this very point is part of my research, however.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

T3 : UNIX, C
dos, C

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

T3 : As a researcher, I don’t find this difficult, so much as I find it a fascinating arena for
experimentation. Nevertheless, in many senses this is perhaps the most difficult bit, since I tend to work in areas where there is endless possibility for this mapping.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

T9: I don’t tend to find this as crucial as 5.1, but then again that’s probably because the design of this usually follows fairly directly from it.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

T9:

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

T9: My considered view is that, interesting as the space of possibilities are here, time spent on other matters tends to be far more fruitful than endless tuning of the parameters and components. So I tend to use fixed overall choices for these, subject to change at whim, or following recent results found in the literature. Operators are really a different matter from the other things in these two questions though; it’s within these that you can stick appropriate domain specific knowledge, or hybridise with other ways to solve the problem.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

T9: Defining one or more distance metrics between genotypes and/or phenotypes is often appropriate for various reasons. Also, my GAs tend be parametrisable to SA too, so components are required which enable me to view acceptance rates at different temperatures, so as to establish a good initial temperature.
*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:*

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?*

*Ty* : Usually there are benchmark results available; if not, then always compare with SA and various kinds of hillclimbing.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?*

*Ty* : Not sure I understand this, although I tend to favour the production of multiple distinct solutions. This is mainly where the distance metrics come in. The results are better to the extent that there are multiple solutions with a good average distance between them.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?*

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.*

*Ty* : You can see what’s going on! In some cases of course – like photofit generation, or evolving art, this is necessary anyway.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.*

*Ty* : You can see what’s going on! In some cases of course – like photofit generation, or evolving art, this is necessary anyway.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.*

*Ty* : Very flexible, to the extent that the user requirements can be varied.
*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

T9: Could be easy to hide what's really happening.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

T9: People tend to like this simply because it shows there's something actually happening. If a 'when fitness gets here we're fine' line is on the graph, possible for most problems, then the illusion of understanding the GA's progress is comfortably strong.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

T9: None really, except there might be much messing around on some problems to translate fitnesses into graphable quantities – eg: if half the pop is between 0.00001 and 0.00002, with others around 21,345,789,329.6

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

T9: In one sense, this doesn't tell you much more than the and visualisation of the entire pool (or bits of it), provided you already know what the selection pressure roughly is. It is occasionally interesting when a very lowly fit chromo is chosen for parenting, but you already know that will happen every now and then. What's interesting, though, is when poor parents lead to strong children. I had some success in a program which indicated on screen every example of a crossover
yielding a child better than both parents. Among other things, this is good to help users justify the use of a GA to their bosses.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

T9:  

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

T9: Entertaining

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

T9: You need to be very selective, since there are so many.


T9: Entertaining


T9: Surely gets very dull after a while, so maybe use only as an option for demos, new users, etc.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

T9: As one of potentially many such diversity measures, this is certainly an important thing to
show. Helps much in seeing what's going on

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

T9: Needs more sophistication to be truly useful. Eg, there may be several best-fit chromosomes, all genotypically or phenotypically distinct.

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

T9: Generally, info on a variety of interesting events – eg: every time a new best fit arrives, let's see its parent(s) and the operation which produced it. Let's also see those operations in which very good parent(s) led to terrible children. Good also to see what's occurring in and between niches. Eg: an ongoing measure of how child fitness correlates with parent diversity.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

T9: Very useful.

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

T9: Very useful.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

T9: An intriguing but strange idea; like getting Fred next door to do brain surgery on you by trial
and error with a soldering iron.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

T7: Pedagogically nice, I suppose.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

T7: The ability to reinitialise the population in any of various ways at one's chosen time. Altering things like penalties for the cost function.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

T7: On GAs in general, I have too many comments to give and not enough time. A good practical thing about making them easier to use – assuming we’re considering a typical industrial setting – is an on-screen estimation, probably dynamic, on how long it will take to reach a given desired fitness. A large scale approximation based on fitness graph gradients would be fine.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

T7: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.
Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.
Respondent - R1

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
R1: about 2 months

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
R1: playing Prisoner's Dilemma

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
R1: I am a computer science student at Cambridge University, and for my project I am working with relating GAs to visual images of creatures, kind of like Todd and Latham's stuff.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
R1: 486 SX 25 MHz : MsDOS, Turbo C++
Solaris V UNIX gcc compiler

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
R1: in the case of Prisoner's Dilemma this is really easy. I imagine that the mapping onto geometric objects will be much harder
*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

*R1: Again, for PD this is quite easy. The problem I am having is that I want to set up a test for diversity, and I don't know much about statistics so I have had to make it up as I go along, and I'm not convinced that it's very good.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

*R1: Initial population creation is purely random, so no problem there. Selecting parents is where I have a lot of problems. At the moment I have it so that about the best 10% produce offspring, mating with randomly chosen partners, but I am finding it a pain to come up with a way to make the number of offspring proportional to score - particularly since I have a fixed population size.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

*R1: This is the biggest problem for me. Population size I have figured out by experimentation needs to be around 100, otherwise nothing good ever develops. Finding a good mutation rate is a nightmare. At the moment I have a mutation rate changing on the fly, to try to help wipe out large populations of the same thing.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

*R1:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?
$R1$: The chromosomes play 10 games, each game being 100 rounds against randomly selected opponents from the population, the points add up to give a fitness measure. This is then multiplied by a measure of diversity giving an overall score.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

$R1$: using the diversity test - here I compare the values of the genes in each chromosome with the average value for each gene and the higher the difference, the greater the score.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

$R1$: Massive amount of information!

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

$R1$: Too much information for a human to usefully digest.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

$R1$: Limit the number of things the user has to examine.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

$R1$: The user may not know what he/she is doing, and could pick an non-representative selection, and miss the interesting things.
*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

R1: See how quickly the chromosomes converge on a solution, also see how stable populations are.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

R1: only useful in conjunction with the other graphs showing fitness, otherwise for example, you could get graphs which bottom out after n generations, but don't tell you how well the populations were actually scoring, just the fact that they had reached a stable state.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

R1: 

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

R1: Way too confusing - I don't see how a user could get any useful information out of such a picture. Perhaps I'm wrong. Actually, I guess that in the testing stages, when the population size could be kept small, it would be reasonable.

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

R1: 

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

R1:


R1: Would be interesting, I guess!


R1:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

R1: Aha! This is very good. I think it is necessary in populations that tend to stabilize, in order to keep variation going (if that's what you want) but if you just want to solve a problem that could also be useful to push away from local maxima.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

R1: As I said above, I don't really know how to implement this sort of thing - it sounds like statistics to me. Could be quite computationally expensive, maybe?

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT
YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

\textit{R1:}

\textit{*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTER-
ACTION OPPORTUNITIES FOR YOUR USE OF GAs?}

\textit{*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN,
PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECU-
TION:}

\textit{R1:} very useful - like an omniscient, but impotent viewer.

\textit{*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:}

\textit{R1:} Could cause problems, but I think it would be really interesting, as the user could get the
chromosomes away from local maxima, which is exactly the sort of thing humans are good at. Also
simulates a kind of real environment, which changes over time, and could test chromosomes ability to
adapt in a changing environment, maybe.

\textit{*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:}

\textit{R1:} Why? this seems silly and only useful for initial testing of the program. I must be missing the
point.

\textit{*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A
GENERATION:}

\textit{R1:} Ditto.

\textit{*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD
CONSIDER BENEFICIAL.}

\textit{R1:} Changing population size, and even chromosome size might be useful.
*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

*RI*: I hope that my project will end up with a way to visually display chromosomes (ie the phenotype) and see how fitness of the genes to solve on sort of problem relates to their appearance. Perhaps I’ll let you know if I get any interesting results!

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SERIES?

*RI*: no
GA Visualization, Design Questionnaire.
Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.
Respondent - R2

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
R2: 1 year

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
R2: Research, mainly in neural net construction

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
R2: Seemed like an interesting idea at the time ;-)  

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
R2: C++, using custom-written graphics to display network weight values, and errors, etc.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
R2: Only a problem when there are lots of constraints

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
R2: Difficult when the final fitness is a function of a number of attributes. E.g., if you want to minimise cost, while maximizing productivity, while...
Also, when the fitness landscape is very flat, apart from a few localised peaks. A while ago I tried to get a GA to come-up with a XOR circuit, using OR, AND, and NOT (I think?). Anyway, solutions which were very close a solution were really unfit. I gave up trying to design an effective evaluation function.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?  
R2: Creating initial population can sometimes be time-consuming when there are a number of constraints. I would then try to design an evaluation function which would not violate these constraints, given two 'legal' strings. The XOR problem above is a good example.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?  
R2: Creating initial population can sometimes be time-consuming when there are a number of constraints. I would then try to design an evaluation function which would not violate these constraints, given two 'legal' strings. The XOR problem above is a good example.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFiculties YOU ENCOUNTER BELOW.  
R2: 

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:  

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?  
R2: Just by assessing the fitness functions
*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

R2:

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

R2: If clusters were forming around local minima

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

R2: Depends on the population size I suppose. Execution speed might be a problem, e.g. for a 'director' who wished to observed how the population changed overtime. Slow updates might make it more difficult (from a cognitive) perspective to observe this.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

R2: as its user-defined I can't foresee any problems PROVIDED the user know what subset of strings he/she wants, and how to specify them

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

R2:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
R2: Good to see if population is stagnating, and might need a boost (e.g. load of mutation, or a few new randomw strings)

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
R2: can't think of any

*Q.8. AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
R2: probably not a) the easiest thing to do in general b) the most useful but probably is problem-dependent to a large degree

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
R2: could get tricky!

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.
R2: dunno

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.
R2: dunno

*Q.8.3.A: THE INTERNAL ACTIONS OF THE GENETIC OPERATORS BEING APPLIED
TO THE CHROMOSOMES, E.G. THE SPLITTING AND CROSSOVER BETWEEN TWO CHROMOSOMES BY A SINGLE POINT CROSSOVER OPERATOR - ADVANTAGES.

*R2*: Probably good for education purposes, but when you think of the number of matings that are going to occur in a typical GA run, then it'd probably be too much to take in.

HOWEVER, if your mating function were in some way 'intelligent', then you, the director, may want to observe how the crossover (or selection process) was being decided/ performed.

For example, you might have a selection process which (instead of the traditional, biased roulette wheel) involved the strings wandering around a grid until they find a mate they fancy and then reproduce. The 'attraction' function might be an evolving AI module of somekind, and you mind want to observe it working.


*R2*: see above

*Q.8.4.A*: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOCIs”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

*R2*: good to show clustering around peaks

*Q.8.4.D*: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOCIs”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

*R2*: 
*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R2: It might be interesting if you could look at a 'family tree' of an individual chromosome, and see how the fitness improves.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R2: extremely useful

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

R2: e. useful

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

R2: e.e. useful

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

R2: e.e.useful++

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R2: an AI module which monitors the 'directors' behaviour and learns how to direct the GA itself (only joking (well 75% joking!!) )
*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

_R2_: Saving a GA run, and replaying it at a later date? Sorry, I’ve exhausted myself Trevor... Hope this is of interest.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

_R2_: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - R3

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

R3: Nearly 2 years.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

R3:

Optimization of systems
- standard test suite for GA’s
- real world problems (greenhouse control, satellite movement)
- dynamic optimization problems

My first task was the development of a Genetic Algorithm Toolbox for Matlab (during my time in Sheffield one year ago). This toolbox is available from me.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

R3: The implementation of GA’s is straightforward. They are powerful. Using, for instance, gradient based methods, is often not possible for real world problems.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

R3: Matlab - on different computer systems (PC and SUN Sparc). If you don’t know Matlab: this is a powerful programming and visualization environment available on nearly every computing platform. The development time for a system is short, because of the huge number of problem
specific toolboxes. Especially in control Matlab is widely used. (The drawback: Matlab is expensive. However, most university own site licenses - have a look to the control group).

Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
R3: No real problem. The Toolbox can use real and binary variables. Until now, all of my problems used real parameters. However, I know, that there are a lot of problems, were the mapping/embedding is difficult.

Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
R3: Here goes the work. 80%-90% of the time for programming/solving the problem is needed for implementing the evaluation function.

Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?
R3: If I don't define special parameters the Toolbox uses default parameter. This includes every part of the algorithm. Thus, if I don't know a lot about the system, I work with the default ones. On the other side, I can change everything. However, most of the time I don't have to.

Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E. G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?
R3: see above

Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU
ENCOUNTER BELOW.

*R3:* Not at the moment. However, I think about implementing something, but I didn't find a clean and general way of doing it. Every system is different. On the other side, quite a few problem need a sophisticated preprocessing. This could speed up the optimization considerably.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

*R3:* Have a look to the data/results. You have to understand the problem, otherwise you can't weight the solution of the GA. Or, when you get results try to understand them - this is often the way to learn more about your system. I didn't find a global way for weighting. The given best solution of the GA depends very much on the evaluation function.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

*R3:* see above. If the evaluation function is smooth, you don't need such a long run of the GA, if it is more chaotic - then you have a problem and you are lost in the GA-space.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

*R3:* if only one generation (the actual one) at once

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

*R3:* too much information, you don't have to see every bit of information.
*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

R3: necessary, if you know/understand the overall meaning of your data, you want to have a look.

For instance, I plot the chromosome of the best individual in every generation over all generations. Thus, I get a meaning of the change of the best individual during the optimization.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

R3:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

R3: If you plot the fitness of the population (only best individual and/or mean and/or worst), this plot includes the gradient. Thus, normally it is not really necessary if you have the fitness directly. However, one of them is absolutely necessary.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

R3:

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

R3: see 8.3
*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

    R3: see 8.3

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

    R3: see 8.3

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

    R3: see 8.3


    R3: If I want to know, how the GA works, this would be useful. but see below.


    R3: This would be much to operator specific. Using GA I want to solve problems. I don’t wanna know how the GA works.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

    R3: Similarity to the fittest sounds like a good idea. (I do an histogramm of the differences between all individuals between each other. This gives a meaning of the diversity of the population.)
*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

R3:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R3:

At the moment I visualize the following things (every 10 generations or so):

- fitness value of best individual in the last 20-40 generation
- chromosome of best individual in the last 30-60 generation
- all chromosomes in the actual generation
- all fitness values in the actual generation
- histogramm of diversity of chromosomes in actual generation (first try, needs more work)

This is quite enough for a good understanding, what’s going on.

(For every system I often include system specific visualizations (dynamic optimization - i.e., results of simulation with best individual for instance.)

There are lots of new possibilities, if you can make movies and so on. At the moment, the computing power is far too less to think about an implementation. However, I think there should a lot be done. I will do some thinking as well and when you contact me, we can talk about more ideas.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?
*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R3: This is/was quite useful for me. During the solution of my first problems I needed such a control panel and thus implemented one in Matlab. If the computing power is high enough or the problem simple, this online control is useful. However, now most of my problems take hours of computing time. Thus, I run the GA offline and save all (intermediary) results.

*Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:

R3: see above. Most of the time you don’t have to. Nevertheless, some problems are easier to solve, when you change parameters during optimization. For this, you have to understand, what’s going on. With my control panel I could change the parameters on the fly, even without breaking/stopping the calculation - should be useful in a control panel.

*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:

R3: Huhh, what are mutation abd recombination and so on for? If the operators are good, you don’t should do this. If not, change your operators or look for a better mapping/embedding of the problem.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

R3: You divide between population and reproduction gene-pool. I am not sure, that I get the difference. My populations are my reproductiongene-pools. Am I missing something?

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R3:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE
EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAS? PLEASE NOTE THEM BELOW.

R3: The implementation of a visualization tool used by many people is quite difficult. If you could define a really portable format for the data...

I would like to hear more about your thoughts.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

R3: I would appreciate being contacted in the future.
GA Visualization, Design Questionnaire.
Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.
Respondent - R4

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
R4 : approx. 2 years.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
R4 : Various optimisation problems, particularly very difficult problems in the real world!

Designing a GA toolkit, GA meter. I have been using GAs because the problem domain can be separated from the search domain, hence generic toolkits are possible.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
R4 : Traditional techniques, if they exist, are extremely computationally expensive to use on the size of problems that I am using. GAs (and other heuristics, such as SA, TS, etc.) seem ideal for these class of problems.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY IE. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
R4 : UNIX machines (various) / PCs
Language: C
Toolkit: GA meter - a generic GA toolkit developed at UEA. It has a user-interface with many of the facilities you mention below. Problems can be integrated into GA meter VERY easily and every parameter can be changed interactively even at run-time.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

*R4: This is usually the most important stage and will more than usual, dictate the ease (or lack of) that the following steps will be implemented. This is, arguably, where the clever thinking is required when using a GA - that or luck.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

*R4: This will wither drop out from the problem objective or the representation. If this is not the case, then this stage may be harder than necessary requiring some subtle technique to return a fitness value.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

*R4: Again, most of these components will be apparent from the representation. For a binary bit string, I use basic operators and see how they perform. For permutation problems, then operators get tricky.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

*R4: I use very basic parameters initially and see if the GA will work using a ’dumb’ GA. As I have said, with GAmeter, it is very easy to change parameters at any time. Thus the initial parameter settings does not worry me too much as I know they can be changed at will.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

*R4: For most of the problems I have tried, No. As mentioned I use a dumb GA. Try it out and
see what results I get. If it is not working, then a re-think of the representation may be required. If it works OKish, then I play around with parameters, possibly new operators, to see if there is any improvement, and how much.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

*R*: Depending on the problem in question. Several alternatives exist...

- Try against an exact method if one exists for quality of solution.
- Try against a specialised heuristic.
- Try against a general heuristic, SA, TS, etc.

The list is endless really. As I said, it will depend on the problem in hand.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

*R*: This is an academic question really. I apply GAs in an industrial context, where the quality of the output is more important than how it fares to all other points in the search space. Hence I usually concentrate on the above step.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

*R*: 
- This is good for seeing how similar (or not) all members of the population pool are.
- It may suggest ways in which improvements could be made to the GA (for example niching)
- It may highlight how the GA has become trapped in a local optimum.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

R4 :

- It can be confusing if you have very large bitstrings. (I sometimes work with bitstrings of '000's bits) - not very informative if the genotype/phenotype map is not straightforward.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

R4 :

- reduces disadvantage #1. (If I understood you correctly!)

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

R4 :

- How do we know the selection is a fair selection?
- Gives another burden to the user to decide.
- doesn't help disadvantage #2.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

R4 :
- Standard visualisation tool everyone can understand.
- Shows convergence of GA, etc.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

R4:

- Too much emphasis can be placed on the graph without going into any detail as to why that pattern occurred.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALISATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

R4:

- Shows how the genetic operators have been working
- Shows which children were automatically discarded.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

R4:

- Is this needed?

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.
**APPENDIX C. GA USER QUESTIONNAIRE RESPONSES**

**R4:**

- Can highlight whether mutation should be increased or decreased.

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

**R4:**

- A lot of overhead for this information.


**R4:**

- can highlight when the operator has become defunct or when another operator would be more useful.


**R4:** as above.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

**R4:**

- could be useful...
*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOC I”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

R4:

- but it really depends on the problem. For example, in highly epistatic problems, it's not the number of different genes but the actual genes that are different which determines how good or bad that solution is. This information would be useless in this case.

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R4: This I am very interested in. As a developer of a toolkit, I am always looking at ways in which the visualisation could be improved. But at the same time I think about the overhead caused by this visualisation.

My outlook is visualisation is nice, but not for the sake of speed and general usefulness. (i.e. it’s no point added some functionality if most problems don’t need this information or visual guidance.)

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R4: All of those options (bar one) are catered for in GAmeter, so I think they are useful! :)

I know why you may want to step backward, but that’s a lot of overhead on the GA.
*Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:

R4: Again this is very useful. Often the GA can be improved if the parameters are adjusted during run-time. (spoken from experience!)

*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:

R4: Useful (and yes, in GAmeter!)

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

R4: Hmmm, I’m not so sure of this one. Since there is an operator which decides which solutions enter the population pool, so you need to edit the reproductive pool? (Especially if you can edit the population pool)

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R4:

- Well, there’s problem specific interaction. For example changing a problems variables - or displaying the solution graphically which you can only do with some problem knowledge there.
- There’s displaying (not really interacting) a series of results from a set of experiments. Useful in seeing how(if) consistent the GA is.

I guess the list is endless, but there is a limit on how useful all these interactions are.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAS COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAS? PLEASE NOTE THEM BELOW.

R4: If you are interested in seeing GAmeter, you are more than welcome to. It is free for academic purposes.
Perhaps it may give you a few more ideas, or more likely, you can suggest future improvements.

GAmeter is continuously evolving and I am always on the lookout for new ideas, etc.

It sounds as if you have already thought about many of the options that are already included.

e-mail me if you are interested... jwm@sys.uea.ac.uk

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

R#: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - R5

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

R5 : Six months (studying them 2 years).

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

R5 : Standard cell placement. It's a small part of the problem of designing silicon chips.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

R5 : My interest in GAs comes first. Somebody suggested the placement problem as a hard optimization problem that might even have some money in it. As to why I like GAs... that's really because as you study the subject, your mind is throwing up ideas for improvements almost as fast as you understand it. That is, the subject is young and scruffy.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

R5 : HP 700 series work stations...

Model 710 712/60 720 735/125
Ram 48Mb 64Mb 64Mb 144Mb
Disk 500Mb 1Gb 500Mb 1Gb

735 is 2-3 times faster than the others.

I'm writing in Ansi C, I'm hoping to move over to C++ "When I've time" to learn it properly. I'm
already trying to use the object oriented philosophy in C. For example, my large number of functions are partitioned into files, one file for each "class". I try to produce functions for as low a class as possible for the sake of reusability. For example, I have "solution" functions, but a solution is a huge ugly thing that is derived (by the measure) from an -ordering- (that is an array of integers 1-n in some order). I've made crossover specific to the ordering class, because that may become useful for other projects.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

R5: That's The Problem. That's what makes the use of GAs like the mensa test. Recognizing what the parental contribution should be, and then figuring out a representation that supports that. I've got a friend whose head of department hired someone to solve a problem using a GA. She seems to have used any-old representation, and now she, and her boss, go round telling people GAs don't work. (Don't worry, I've offered to help out if they send me more info.)

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

R5: For this project, yes, that took time. But that was "just programming". I've made at least one very silly mistake that wasted a lot of time. Actually, it was the sort of mistake that Object Oriented Design would have made impossible, which is why I'm mending my ways.

Also I think I could have made more use of existing packages in the electronics field. But so far I've had to work without input from electronics professionals, and my publication (pending) will state that I am addressing a simplification of the industrial problems. You'll be glad to know I'm about to start a PhD at York - in an electronics department.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION
METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

*R5: Difficult isn't the word... my whole approach is based on a special kind of population seeding! i.e. population creation. As for the other GA components you mention, really I've stuck to somebody else's published details about an algorithm, which I am trying to improve upon.

*Q.5.4: SELECTING PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

*R5: This is a real crusher, this is where your package would save a lot of time. Setting parameters is an agony for me. Every time I run the thing it takes more than a day, at the end of which all I know is that the run didn't work. It would be nice to be able to watch the run and monitor population diversity, and population movement. To some extent, the setting of parameters is -irreducibly- hard. There are theoretical methods for setting them, which work when you know a lot about the problem i.e. it is a toy problem.

I don't know if it's possible, but I'd like to see your package built with the idea of users contributing add-on modules. However thoroughly you build the thing, when I use it I am going to want to add more, and I would want to send my modules to some centre of cooperation.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

*R5: My approach is to use an assisting optimizer to produce a paradigm solution which is partially optimized. Then I produce a population from it by scrambling the solution with small changes to the gene values (the ordering problem is a high alphabet problem with a metric i.e. some of the alleles are "near" one another.) The aim is to concentrate search on a small part of the solution space which is yet expected to contain global optimum. So far the assisting optimizer is just another GA, but in other domains it might be a different optimizer.
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

Say the word and I will send you my paper. I presented it at the recent AISB conference on Evolutionary Computation.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?
R5: I'm just comparing solution quality to that found by my rival's GA.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?
R5: I'm running the GA repeatedly from different starting points (the random string optimized to produce the paradigm string). Then I'm comparing GA outputs in terms of fitness, absolute phenotype features, and phenotypic features considered more abstractly. I believe that my problem has a large number of global optima, that are the same in statistical profile but very different in how that profile is instantiated.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.
R5:

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.
R5: If you can understand it, your problem is too simple.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - AD-
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

VANTAGES.

*R5: Establishes conventions.

*Q 7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

*R5: We do not yet know enough to establish such conventions, yet maybe your package should make a stand and be open to change. At the moment everybody does their own thing here.

*Q 7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, i.e. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.


*Q 7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, i.e. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

*R5: 

*Q 8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q 8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

*R5: Could be used to spot convergence.

*Q 8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

*R5: Your package should spot convergence for us. Get rid.

*Q 8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - ADVANTAGES.

*R5 : None.

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

*R5 : Distraction. User should know where mutation will occur.


*R5 : Good for educational purposes. Also, all of your ideas for displays might turn out to be useful for debugging the GA.


*R5 : None as long as it's optional.

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

*R5 : This justifies your enterprise. If I don't know this I don't know what my algorithm is doing (that premise currently true!)

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.
R5:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R5: Niches. I mean sort strings so that similar ones are together.

By the way, I'd like string similarity to be pretty flexible, or at least particularly open to user add-ons.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R5: Excellent.

*Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:

R5: Definitely a good idea.

*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:

R5: A creative idea, yes I’d like to try that (though I hesitated a moment). Yes, real biologists as well as observing, they do experiments like stealing a lion’s cubs to observe the reaction. We should certainly be able to do that.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

R5: Personally I cannot see myself bothering with that. I can’t see that helping to study -macroscopic- population behaviour, which is the important thing.
*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R5: Can't think of any others.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

R5: The above covers it I think.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

R5: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,

The Open University, Milton Keynes MK7 6AA.

Respondent - R6

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

R6: 2 years.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

R6: Determining the best transmission scheme and data rate for a baseband communications system. Designing FIR filters. Producing bit sequences with special autocorrelation functions.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

R6: It was more a case of selecting tasks that the GA could be applied to - I’m working on a project to investigate the use of GAs in the design of communication systems.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

R6: A network of Sparc stations, running UNIX, with self-written software (written in C++, using Sun’s compiler).

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

R6: Nothing - this is generally very straightforward.
*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

R6: It is important to be careful that there are no weaknesses in the evaluation function definition, as the GA has been seen to exploit them. Producing effective evaluation functions is most difficult when a trade-off or compromise is required between a number of system performance measures.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

R6: Early experiments produced a reliable structure for the GA which has been applied without any problems to a variety of applications.

*Q.5.4: SELECTING PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

R6: See previous comment.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

R6: No.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

R6: In some of the cases, the ideal solution is known. The GA has been found to produce close to ideal solutions.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

R6: Haven’t bothered.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.
R6: Variation between members could be easily observed. Convergence could also be spotted easily.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.
R6: Too much information displayed at once could hide useful information.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.
R6: Less of an 'information swamp'.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.
R6: How does the user define a 'representative' set of chromosomes. They may well NOT be representative at many or all points of a particular run.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
R6: Gives an indication of convergence.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
R6: The fitness vs. generation graph is usually very noisy, particularly with high variations in
the fitness function between good and bad members. The gradient of this curve would have to be averaged to produce a useful value, and the averaging needed may well depend on the particular application.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

*R6*: For binary strings, it would be possible to determine whether every possible allele existed in the initial population, and how well different alleles propagated.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

*R6*: For non-binary strings, the number of possible alleles for each gene is likely to be prohibitively high.

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

*R6*: This could give an indication of whether the mutation rate was too high (interfering with the evolution process) or too low (allowing stagnation). It could also indicate the introduction of a previously unencountered allele on the chromosome.

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

*R6*: None.

*Q.8.3.A: THE INTERNAL ACTIONS OF THE GENETIC OPERATORS BEING APPLIED
TO THE CHROMOSOMES, E.G. THE SPLITTING AND CROSSOVER BETWEEN TWO
CHROMOSOMES BY A SINGLE POINT CROSSOVER OPERATOR - ADVANTAGES.

*R6: Possible to observe correct operation of the GA.

*Q.8.3.D: THE INTERNAL ACTIONS OF THE GENETIC OPERATORS BEING APPLIED
TO THE CHROMOSOMES, E.G. THE SPLITTING AND CROSSOVER BETWEEN TWO
CHROMOSOMES BY A SINGLE POINT CROSSOVER OPERATOR - DISADVANTAGES.

*R6: Unnecessary.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE
THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHRO-
MOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT’S BIT
POSITIONS (“LOCT”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

*R6: A better indication of convergence than the gradient of the fitness vs. generation graph.

*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE
THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHRO-
MOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT’S BIT
POSITIONS (“LOCT”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

*R6: None.

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT
YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

*R6:

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTER-
ACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN,
PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

R6: Useful - if the GA is not converging, altering the mutation rate could help.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

R6: For my use - limited.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

R6: I use steady-state GAs, so the gene-pool is the same as the population.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R6:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

R6:

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

R6: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - R7

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

R7: AI course Fall 1991 w/Melanie Mitchell. Fascinating. For research since 4/94.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

R7: Computer Architecture/Microprocessor Design.

This is a multi-dimensional combinatorial optimization problem with multiple objectives.

The problem is this: How can I partition millions of transistors into dozens of on-chip hardware structures (memories, adders, etc) to satisfy multiple budget constraints and multiple objectives toward identifying a set of near-optimal partitions? Paper in upcoming 6th ICGA conference.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

R7: I confess that I have always found it fascinating and jumped at the justified chance to play with it in my research.

In addition, it has several characteristics that make it appropriate for my specific design problem. Among these characteristics:

+ It is readily parallelized on networks of engineering workstations. This is how real-life design engineers work.

+ My objective function is very long (5+ hours) and I need a parallel approach.
As a designer, I am looking for sets of near-optimal solutions with which to study and propose the next design improvement. What are the design and performance characteristics of near-optimal solutions, and what should I do next to improve the design further? I am less interested in a single point in the design space.

The GA readily handles multiple objectives.

The GA is a true global search technique.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY IE. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

R7: Language: C for my GA and objective function/simulators.

Heterogeneous networks of workstations running variants of UNIX.

Primarily, I use DEC workstations (MIPS-based) running ULTRIX. Sun workstations running Sun-OS. I may get involved with HP workstations too, and have already ported the code over to HPs. I use DECs for development, and Suns for full-blown runs.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

R7: For my problem, this is not too much of a problem. For other problems, this is a major issue, indeed perhaps the single major problem leading to success or failure of the GA in a non-traditional representation domain.
*Q.5.2: PRODUCING AN EFFECTIVE FUNCTION?

R7: I call a simulator in my objective function. It took me a long time to write this simulator. However, this is independent of the GA; the simulator is hard whether I do hand-optimization or any other global technique. There is nothing GA-specific about the simulator.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

R7: I read the literature and integrated what I learned. I admit this may not be optimal, and knowing what is optimal would be good.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

R7: See 5.3

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

R7: Nothing special. I keep 200 networked workstations busy in parallel. So I take great care to make sure everything is debugged and in-order before I run. But, that is the nature of research, eh?

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

R7: Repeatability is a major way for me to know the GA is not simply hacking about. If I can get *nearly* the same design multiple times, I am confident that the stochastic optimization is OK. Also, I sanity check, and plot the gene and objective values as the simulations proceed. Premature convergence has been the biggest GA problem I have had to address. Once that was solved, things
seem to be pretty good.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

R7: See 6.1

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

R7: Fun. Good to teach non-believers. Could get an early-on sense of what was happening. I do this already by making plots and it is semi-automated. As such, I expect that other serious GA researchers do the same.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

R7: I can imagine "visualizing" real-valued parameters. But how to you visualize other problem-specific genetic representations, e.g., tree-based, etc.? Is a general-purpose display method even possible in consideration of the number of possible genetic representations?

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

R7:

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

R7:
Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, i.e. THE
GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
R7: Yes since this tells the user s/he may be nearly done, or at least near a plateau. Also, I watch
the standard deviation of objective values.

Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, i.e. THE
GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
R7:

Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION
COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED
FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE
USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR
DISADVANTAGES, IF ANY, COULD YOU FORESEE?

Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
R7: Good. I do this already in a non-fancy semi-automated way so I can watch what design is evol-
ing as it proceeds. My entire GA run might take 2+ weeks so I need to watch for sanity as it proceeds.

Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
R7:

Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - ADVANTAGES.
R7:

Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - DISADVANTAGES.
R7:

R7: Fun, good teaching/debugging tool. I am not sure how to digest and exploit this info, however, over 1000s of runs in a real application.


R7:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

R7: Yes, many.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

R7:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R7:

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTER-
ACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R7: Again, fun, and good for teaching but that is all.

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

R7: Again, fun, and good for teaching. I suppose it could be used "heuristically" to change direction of search, but this is an ad-hoc approach to an already stochastic optimization technique.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

R7: Again, fun, and good for teaching. However, probably disruptive to the GA. Hard to say.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

R7: Again, fun, and good for teaching. However, probably disruptive to the GA. Hard to say.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R7:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

R7:

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE
RESULTING GA VISUALIZATION SYSTEMS?

R7: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - R8

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?

R8: 4 or 5 years.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?

R8: Algorithm optimisation, curve fitting. Have also done basic work on GAs themselves.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?

R8: GAs offer a general method for solving optimisation problems; also because of interest in the GAs themselves.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY i.e. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

R8: Various - hard to be specific.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

R8: One of the hardest problems, if not the hardest.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

R8: Hard for combinatorial problems rather than function optimisation; often depends on 5.1 in
such cases.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?  
R8: The operators depend very heavily on the representation.  
Initial population is fairly unimportant. 
Reproduction method is somewhat important, but not the biggest issue.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?  
R8: Not too hard. The trouble is that sometimes, larger populations were needed than could realistically be used...

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.  
R8: Choosing when to halt the GA is another problem.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:  

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?  
R8: In most cases, the fitness function itself was used; in the curve case, visual inspection was also useful.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?  
R8: I hope it isn’t! I just want it to explore the “right part” of the space. More seriously, multiple
runs were used to see how often the same solutions resulted.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

R8:

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

R8: Far too many for this to be useful.

The chromosomes themselves are not very meaningful in some of our work - i.e., complex calculations are needed to derive the individuals from them.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

R8:

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

R8: Same.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

R8:

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
Not a lot, because the fitness versus generation graph is quite noisy, and so derivatives would be measuring noise.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

*R8: *

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

*R8: See above.*

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

*R8: *

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

*R8: See above.*


*R8: *

R8: See above.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE OF ITS BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

R8:

*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE OF ITS BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

R8: Far too simple a measure of similarity.

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

R8:

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAS?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

R8: Useful for debugging, but not in production runs.
*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

R8: Maybe useful if the GA gets stuck.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

R8: See earlier remarks.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

R8: See earlier remarks.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

R8: Ability to introduce a strong mutation pulse to kick the GA into a different region of solution space.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

R8: The really tricky issue is designing the representation and the operators, not controlling or visualizing the GA during running.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

R8: No. I would not object to being contacted in the future.
**APPENDIX C. GA USER QUESTIONNAIRE RESPONSES**

GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A1

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?*

A1: For almost a year.

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?*

A1: Optimization of designs in the mechanical engineering field (and optimization of some test problems).

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?*

A1: Previously, a Monte Carlo method was used. We found out the GA was much more effective at solving large problems.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?*

A1: We wrote our own GA programs. Based on the first Fortran program on an MS-DOS compatible that used the Monte Carlo method, a GA was developed for that environment. Later on, a version was written in C. It was a more powerful and flexible version and it was developed on an Amiga. It is written in ANSI C and is therefore portable.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?*
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

A1: When solving mechanical engineering "optimization" problems, the strings are the collection of independent design variables, so this mapping is no problem at all.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

A1: That can be quite difficult. It would go too far to go into detail about the problems that can arise, but there are problems in the field of multi-variable fitness evaluation. Furthermore, in the field of mechanical engineering, criteria are often determined by the "contractor" and the importance of these criteria are not as "fixed" as you would like.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

A1: The ANSI C version of the program we developed can use multiple mutation and crossover methods. These can be selected at runtime and turned on and off while the algorithm is running. This gives the user the ability to experiment with the different methods and to gain more insight into them.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

A1: Population size is chosen rather arbitrarily at the moment (limited by memory and practical speed limitations). Mutation rate as well as crossover rate can be adjusted at runtime (constantly).

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

A1: We did design some preprocessors to make problem definitions easier, but these have nothing to do with the GA. So the answer is: no.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:
*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

AI: In mechanical engineering, it is quite easy to check the results against existing designs. We have done this in some cases.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

AI: We didn't do extensive research on this, because we got satisfying results, which indicated that the problem space was searched quite well.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

AI: A detailed "report" of the current population, providing a lot of data to those that can "read" it correctly.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

AI: Might confuse people, and might tempt people to draw the wrong conclusions. I mean, the GA is strongly stochastic, so one must always be careful about drawing conclusions from any particular run.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

AI: Would give even better controlled data. This would definitely be better than 7.1.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.
AI: Same disadvantages, with the additional risk of accidentally disregarding data that might be important after all.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

AI: Not sure about these...

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

AI: Not sure about these...

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

AI:

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

AI:

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

AI:

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

AI:


AI:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCT") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

AI: All these statistics would give more insight into the GA, and would therefore be quite nice for educational purposes.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCT") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

AI: It slows down the GA and therefore (in my particular application) the optimization process, which is a disadvantage (that might or might not be important, depending on your goals, computer speed, etc.).

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.
In the field of optimization, the independent variables are also displayed (at least, we display them). In general, one could perhaps say that the "data that the chromosomes represent" should be shown (optionally).

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:
AI: We have already implemented "start" and "stop" and interactive changing of the parameters.

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:
AI: This can be very useful, to speed up the algorithm and to increase the user's insight into the process.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:
AI: I don't really see the use of this. But perhaps I'm missing something here. It wouldn't hurt as an option, that's for sure.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:
AI: Same as 10.3, don't see the use at the moment.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.
AI: Nothing comes to mind at the moment.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE
EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAS? PLEASE NOTE THEM BELOW.

A1: I am still learning more about GA's every day. In the field of mechanical engineering, we have used them to optimize designs. I keep talking about "we", and I think I should explain myself. I am studying mechanical engineering at the Delft University of Technology and "we" refers to my professor and other people at the department of mechanical engineering design.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?
A1: No. I would not object to being contacted in the future.

I'd be happy to further discuss aspects of GA's and the visualization of them. I can send screenshots of my program (that is the ANSI C program, which I wrote) that show how I visualized everything. Of course, this is only an example. I know there is room for improvement. I am therefore interested in any new suggestions you might have on this subject.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A2

*Q:1: HOW LONG HAVE YOU BEEN USING GAs?
A2: 18 months

*Q:2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A2: I am using GAs for the design of Predictive Controllers.

*Q:3: WHY DID YOU USE GAs FOR THIS TASK?
A2: Because classical methods of optimization cannot solve the problem mentioned above.

*Q:4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A2: Currently I am programming in Borland C++ v.4 for DOS. After next month I will start to use Linux OS and gcc.

*Q:5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q:5:1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
A2: Not difficult

*Q:5:2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
A2: Not difficult
*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

A2: Not difficult but complicated

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

A2: complicated

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

A2: You have mentioned everything

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

A2:

1. Extend searching
2. Quality of the solutions (final performance)

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

A2: The main point is the final performance considering at the same time the current practical issues.

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRE-
SENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

A2: Supervisory control to all the individuals.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

A2:

1. For large populations?
2. It’s difficult to check all the candidates

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

A2: Ability for someone to be experimented, so that to choose the best possible representation for the specific problem.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

A2:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

A2:

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

A2:
*Q.8: As well as directly illustrating the output of the GA, visualization could be used to represent additional information either derived from the output dataset or recorded separately. If visualization were used to represent the following characteristics what advantages or disadvantages, if any, could you foresee?

A2:

A2:

*Q.8.2.A: The occurrence of mutation in chromosomes where a mutation operator has been applied - Advantages.
A2:

*Q.8.2.D: The occurrence of mutation in chromosomes where a mutation operator has been applied - Disadvantages.
A2:

*Q.8.3.A: The internal actions of the genetic operators being applied to the chromosomes, e.g. the splitting and crossover between two chromosomes by a single point crossover operator - Advantages.
A2:

*Q.8.3.D: The internal actions of the genetic operators being applied to the chromosomes, e.g. the splitting and crossover between two chromosomes by a single point crossover operator - Disadvantages.
A2:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.
A2:

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.
A2:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.
A2 : You have not left anything for me to think. The questions mentioned above, definitely have only advantages.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:
A2: HELPFUL

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:
A2: HELPFUL
*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:
A2: HELPFUL

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:
A2: HELPFUL

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.
A2: You have thought everything.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.
A2: The transfer of the biological terminology to GAs field must be more direct and more conceivable.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?
A2: Yes. I would object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A3

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A3: two years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A3: for biological applications: aligning protein sequences, folding RNA molecules, finding the best set of parameters for a specific application.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A3: the problem of multiple sequence alignment with the 'sums of pairs' used as an objective function is known to be NP-complete. Thus, as the problem can be approached in a combinatorial way, it looked like a good idea.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A3: UNIX, C & VMS, C

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
A3: I found quite generally that the 'naive' approach rarely works. Thus, the mapping seems to me the most crucial point in the strategy of designing a GA.
*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
A3: In my case, the evaluation function already exists, so most of the time there is no real choice.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?
A3: I found that a lots of changes in the selection scheme, generation production and other have drastic effects. But again, this is very difficult to control. I am now working with a model using most of the features described by DAVIS in 'The handbook of GA'.

This is not necessarily the best model, but it works reasonably well, and because of the fuzziness around these parameter I am less and less keen on playing with them. It seems to me much more worth spending time on the quality of the mapping and the quality of the operators.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?
A3: It seems to me that the population size is not a real problem. In my experience, GA are quite robusts regarding this parameter. The population size may be GA/problem specific, but for a given class of problem in a given GA, using always the same pop size does not seem to be a problem.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.
A3:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:
*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?
A3: We use as a benchmark, an exhaustive program that can provide a guaranteed optimal solution for a small problem.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?
A3:

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.
A3: none

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.
A3:

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.
A3: if properly done, it could help visualising the emergence of some niche, and maybe their relations

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.
A3:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
APPENDIX C. GA USER QUESTIONNAIRE RESPONSES

A3:

*RQ.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE
GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
A3:

*RQ.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZA-
TION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED
FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE
USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR
DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*RQ.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
A3:

*RQ.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
A3:

*RQ.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - ADVANTAGES.
A3: allow the the user to really get a feeling of what the mutation does, and possibly, what are its
limits.

*RQ.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - DISADVANTAGES.
A3:

*RQ.8.3.A: THE INTERNAL ACTIONS OF THE GENETIC OPERATORS BEING APPLIED
TO THE CHROMOSOMES, E.G. THE SPLITTING AND CROSSOVER BETWEEN TWO
CHRamosomes by a single point crossover operator - advantages.

A3:

*Q.8.3.D: The internal actions of the genetic operators being applied to the chromosomes, e.g., the splitting and crossover between two chromosomes by a single point crossover operator - disadvantages.

A3:

*Q.8.4.A: A “similarity” rating for each chromosome based on how little they differed to the fittest chromosome, e.g., a ten bit binary chromosome that differed from the fittest chromosome in three if its bit positions (“loci”) may have a similarity rating of 0.7 - advantages.

A3: this might be quite useful in helping to identify problems that need a ‘niche’ approach

*Q.8.4.D: A “similarity” rating for each chromosome based on how little they differed to the fittest chromosome, e.g., a ten bit binary chromosome that differed from the fittest chromosome in three if its bit positions (“loci”) may have a similarity rating of 0.7 - disadvantages.

A3:

*Q.9: Please specify any other direct or indirect characteristics that you would be interested in seeing visualized.

A3:

*Q.10: How helpful, or destructive, would you find the following interaction opportunities for your use of GAs?

*Q.10.1: execution control through the use of a control panel to run, pause step forward, step backward, save a snapshot, and/or stop execu-
**Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:**

* A3: This probably depends on the type of problems to solve. It would probably help for large problems that only need to be solved once.

**Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:**

* A3: no

**Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:**

* A3: no

**Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.**

* A3:

**Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.**

* A3:

**Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?**

* A3: yes
GA Visualization, Design Questionnaire.
Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.
Respondent - A4

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A4: 2-3 years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A4: as a search algorithm and as a gave development system

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A4: interest, evidence that GA work well as search algorithms

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A4: Linux boxes, SGI, Cray supercomputers, all in C/C++ (GNU)

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
A4: Doing analysis of proteins, very simple mapping

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
A4: Hard to describe what a "GOOD" protein is, but that's a problem with the field not with GA
**APPENDIX C. GA USER QUESTIONNAIRE RESPONSES**

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?*

*Af.*:

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?*

*Af.*: Hard to find good parameters, did mostly trial and error, still searching for good parameters

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.*

*Af.*:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:*

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?*

*Af.*:

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?*

*Af.*:

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?*

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION-ADVANTAGES.*
A4:

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

A4:

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

A4:

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

A4: variability of fitness criterion (may be an advantage truthfully)

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

A4:

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

A4:

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

A4: quick analysis of relationships genes

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.

A4: again gives you an overall picture of gene changes

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

A4:


A4: Very interesting, good way to actually see if what you plan is actually working


A4:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

A4:
*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS ("LOC") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

A4:

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

A4:

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

A4:

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

A4:

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

A4:

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

A4:

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD
CONSIDER BENEFICIAL.

A4:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAS COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAS? PLEASE NOTE THEM BELOW.

A4:

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

A4: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.
Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A5

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A5: since 1992, approx 3 yrs

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A5: optimisation, adaptive search to identify design options, integration with NN.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A5: optimiser, good search tool

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A5: Sun Sparc Stations, C lang, POP11 lang

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
A5:

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
A5:
*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?
A5:

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?
A5: yes

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.
A5:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?
A5: constraint satisfaction

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?
A5: heuristics, otherwise difficult

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.
A5: all info

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.
A5: too much info, unnecessary

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.
A5: helpful

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.
A5: there can be a chance to lose novel chromosome structure.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
A5: OK, can give some idea about convergence

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
A5:

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
A5:

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.  
A5: not much info

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.  
A5:

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.  
A5: not much info

A5:

A5: will not be meaningful in multidimensional problem situation.

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT'S BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.  
A5:
*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS (“LOC”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

A5: that is not representative of binary representation

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

A5: best solution achieved every generation

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

A5: would be very good

*Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:

A5: not very good idea, instead an adaptation scheme can be developed.

*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS.

A5: not good idea, that would interfere in GA’s search strategy.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

A5: not good idea.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD
CONSIDER BENEFICIAL.

A5: initial partial seeding of population with some "good" chromosomes (using domain knowledge).

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

A5: some hybrid approach could be useful, like for few generations if GA can not find any improvement in terms of fitness, then may be hillclimbing can be started from that point or even simulated annealing or tabu search.

I am developing Adaptive Search Manager using Fuzzy Expert Systems, which is expected to extract info from GA search and utilise that info/knowledge for effective search.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

A5: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A6

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A6 : 4 years on and off (last 2 continuously)

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A6 : 1. Process Planning
2. Mechanical Design
3. Mechanical Durability Assessment Test setup procedure

Number 3. is my current topic

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A6 : 1. Extension of previous research of another.
2. Feasibility study leading to full research project by another. Expected that GAs would mimic the method nature uses for design.
3. It was expected that GAs would deal efficiently with the large data sample that exist in simulation testing. Also, since we only have to convert data in a forwards direction, the existing problems with methods in current use would not be encountered.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYS-
TEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?

\textit{A6:} HP workstation Mathworks MATLAB

It is expected that the development will also work using MATLAB on an PC.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

\textit{A6:} The initial creation of the population would produce many unfeasible solutions. Development of the representation method has (hopefully) solved this.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

\textit{A6:} Was initially a problem to define one which gave good solutions enough of an advantage over weaker members, but which did not completely exclude these members.

*Q.5.3: CHOOSING THE GA’s COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

\textit{A6:} Initial population (see 5.1 above)

Operators etc. Choice not really a problem. The parameters used with them, however, make a large difference to results and solution time.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?

\textit{A6:} See 5.3
*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING
THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU
ENCOUNTER BELOW.

A6: No GA related ones.

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU
TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

A6: The evaluation of my problem gives a percentage match.

Therefore, a match of 100% is a perfect solution.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF
ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

A6:

*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRE-
SENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION -
ADVANTAGES.

A6: Good for investigation as to what the GA is doing.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION -
DISADVANTAGES.

A6: Far too much data for me, since I am looking at an application rather then the GA itself.
*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

A6: Better than 7.1, giving some info on the current generation.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

A6:

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

A6: Gives info on how the search is proceeding.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.

A6:

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.

A6:

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.

A6:

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - ADVANTAGES.

A6: Good for following what has been happening, therefore the understanding of what is going on.

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

A6:


A6: Same as above.


A6:

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

A6: Would allow knowledge of the diversity of the population. This would allow, through experimentation on your problem knowledge as to how the search was likely to proceed.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

A6:
*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

A6: Very helpful (a definite)

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

A6: Could be quite useful for me as I will also be using a GA to identify the mechanical system. If I obtain a 'better' parameter set for my dynamic system whilst the main GA is running it would be useful to be able to introduce this to the current run.

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

A6: It may be useful to be able to input new chromosomes during the run to allow for expert knowledge to be incorporated.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

A6: see above.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

A6:
Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

A6:

Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

A6: no
**APPENDIX C. GA USER QUESTIONNAIRE RESPONSES**

GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A7

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A7: About 4 years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A7: Studying the optimum structure of the Australian sheep breeding industry.

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A7: There are a large number of variable option, all of which interact. For example - the number of rams and ewes held in nucleus groups, the number of years these rams and ewes are used before replacement, the number of rams available for commercial flocks, the number of good quality ewes being promoted into the nucleus, the selection methods used for all of these sheep and the use of artificial insemination and multiple ovulation.

In most of these cases some intermediate value is optimal (e.g. more expensive methods of selection are more accurate, but the cost tends to increase exponentially for only small gains in accuracy), and the optimal values depend on choices for other components in the system. I have selected 17 variable items for use in my main GA. I have used GAs for other specific aspects of the breeding system, but the answers here relate to my main system.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A7: Borland Pascal (not Windows version) on an IBM compatible 486DX 33Mhz. However, I borrow a Pentium whenever possible.
*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?

A7: Most variables I use are continuous, so I have to map them to a series of integers - usually 16 or 32 to avoid too many large genes. This makes the solution 'lumpy', so I sometimes fine tune it in a narrow range after first finding the appropriate range with several broad scale runs. Several of my variables are percentages and these do not map well to powers of 2.

An early problem was the tendency of many strategies to produce impossible results. For example a common problem was the inability of one of the breeding groups to maintain its population because the ewes were moved out before they could produce enough replacements. In other circumstances (e.g. multiple ovulation) the same strategy might be a winner. I finally fixed this by trying to ensure that the genes would produce a legal result. In the above case by setting the gene to determine the number of EXCESS lambs produced after satisfying the minimum requirements, rather than the ACTUAL number of lambs. This requires reordering the calculation, but always gives a valid result.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?

A7: This is by far the biggest job, as the theoretical genetics are very complex (for me anyway). The evaluation section requires about 80kb of code and takes about 1 second per evaluation on my 486DX33. I have no idea how this type of evaluation could be incorporated into a general purpose GA program, except as unit to be compiled with other GA specific units.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?

A7: I worked through Goldberg's book, using his simple GAs in Pascal, then modified the programs
taking into account his comments on potential improvements and any ideas that came to me at the
time, with allowance for my specific problems, but keeping the GA section general enough to apply
to any other problems.

The main change from Goldberg's simple GA is that when I normalize the fitness function, I set
the average value to exactly 1.0, the minimum to zero, and the current maximum to 2.0. This
requires separate linear scaling for those above zero, and those below zero. It avoids the problem of
a few extremely good or extremely bad values skewing the whole distribution. I have been told that
ranking would do this better, but do not know a fast ranking method that would make any further
gains worthwhile.

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE,
THE MUTATION RATE (IF APPROPRIATE), ETC.?

A7: I could not find any really good guidance here, so I have experimented a bit. I found that
the mutation and crossover rates did not make a big difference, so settled on intermediate values
that seemed satisfactory for my main function. Borrowing from nature, I use a circular chromosome
(i.e. there is always an even number of crossovers). I have done some further experiments now
that I have access to a Pentium and am considering options like variable crossover and mutation
rates. At present I usually use 4-10 crossover sites per pair, and have about 20% of the population
as mutants. However, I do not know of any good methods of selecting these other than trial and error.

My usual population size is 300, which I understand is rather large for a 67
bit chromosome. There is only slight improvement in results compared with 100 or 200, but the
smaller populations have definite tendency to sometimes arrive at a suboptimal level, apparently
due to inbreeding and loss of specific bits in the early stages. I am rather sensitive to inbreeding as
it plays a
critical role in my own sheep breeding structure, so I favour large
populations even if it takes longer to get results.
**Q.5.5:** ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.

*Answer:* I have a file of default settings for all the GA settings and those specific to the problem. These defaults can be changed if necessary. Apart from population size, mutation rate and crossover rate, I have options to switch Gray codes on/off and allow or disallow clones (identical chromosomes). I also set an upper limit to the number of generations (normally 120) for when I run a series of tests overnight.

**Q.6:** HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

**Q.6.1**: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?

*Answer:* I normally test any problem 5-10 times, compare it with hill-climbing results and also use my own intuition to try obvious solutions to find out why the GA didn't use them. It is rare for the GA not to find the best possible solution in 10 tries. Hill-climbing can get good solutions, but I have found that it never gets the best solution unless the starting conditions are artificially set with extreme values for some variables. My intuition never gets the best solution, but can be used as a starting point for hill-climbing to reach the best. I also look at the intermediate calculations used in the optimum to check that they make biological sense.

**Q.6.2**: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?

*Answer:* I rely on my intuition and knowledge of the subject to check any solutions that do not appear to be produced by the GA. Usually by editing to create a test subject, then hill-climbing it on the other genes.

**Q.7:** IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?
*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

A7: I did this when first starting simple test functions, and it did help to verifying that my program was doing the right things.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

A7: I quickly stopped looking at each individual as it is too confusing and not informative.

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

A7: My current system shows the 17 gene values for the best 5 individuals. This gives me some idea how things are going, whether they are converging and which genes are still highly variable.

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.

A7: Although the best 5 are always shown I usually only look at the best one (shown in a different colour) and use the current minimum, average and maximum to check how the GA is going.

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, IE. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.

A7: When I do many overnight runs to test the settings I compare them using a graph of fitness versus generation. This show how quickly different settings reach good values, and how close they get to the highest possible value. This is useful because some settings make the best gains early, but seem to run out of variation and fail to reach the maximum that slower settings can reach. I have to do this manually in Excel form test files produced during the run.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, IE. THE
GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
A7: I do not find this useful for examining the actual results (i.e. in terms of my sheep breeding system) except to get some idea whether further increases might be possible if left for more generations.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
A7: This is only useful when I have made changes to the program in this section and need to check that I have not introduced a new bug.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
A7: No need unless the program is not working correctly.

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.
A7: Same as 8.1

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - DISADVANTAGES.
A7: Same as 8.1

A7: Same as 8.1

A7: Same as 8.1

*Q.8.4.A: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.
A7: There might be some value in a convergence value that summarised the whole population so you could check the rate of convergence and decide when no further gains were likely.

*Q.8.4.D: A "SIMILARITY" RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF ITS BIT POSITIONS ("LOCI") MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.
A7: Doing this for individual chromosomes would be confusing.

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT YOU WOULD BE INTERESTED IN SEEING VISUALIZED.
A7: I show either of two 'graphs' during the run. These are done using the standard ASCII block graphic characters in four colours to give about 16 levels of colour/shading, from full red for high values to full blue for the lowest.

One shows the total number of positive bits in each gene over the whole population. This show me which genes have fully converged and which bits still have high variation.
The other 'graph' shows the actual gene values (ranged from minimum to maximum) to show which values are being favoured and which values are dropping out.

These graphs allow me to glance at the screen and decide whether it is worth stopping, or whether it should run a bit longer. The latter also indicates whether some values are still at a high level even if not present in the top 5 shown individually.

*Q.10: HOW HELPFUL, OR Destructive, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

A7: I find it valuable to be able to stop at the end of any generation, then look at the details of the current maximum, step forward, or continue running. I can save the current complete set when paused and often save intermediate stages in important runs. I have never felt any need to step back to a previous generation. I have an option to store the best individual from every generation in a file, so that I can view the whole run and see which genes stabilised early, and which settled down later in the run.

*Q.10.2: EDITING THE ALGORITHM’S PARAMETERS DURING EXECUTION:

A7: I can do this at any time when paused, (and I can only pause between generations), although I rarely do so, except sometimes to lower the population size if I am feeling impatient. I never change the evaluation settings during a run as there is no point (in my case) in running a GA in a changing environment.

*Q.10.3: EDITING THE POPULATION’S CHROMOSOMES BETWEEN TWO GENERATIONS:

A7: I find it very useful to be able to edit the chromosome. This is often done to compare my intuition with the current settings, or to check whether small variations in the current optimum would further improve it. More usually I find out why my intuition would give a worse
answer. In some cases if my graphs indicate that certain values are not being used I can seed the population with an individual with these values and see if can spread these genes in future generations.

I do the editing by using the current best chromosome as a default, then the edited chromosome replaces the current worst individual in the population.

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL’S CHROMOSOMES WITHIN A GENERATION:

A7: I only edit when paused at the end of a generation. I can’t think of any reason to stop during a generation.

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

A7: In order to evaluate why a particular individual is the best (compared with my own ideas) I have a full 50 line screen of data showing intermediate calculations used in the evaluation function. This is essential to determine the effect of minor (and major) changes to the current settings, to show the effect of each gene in the whole picture, and to allow me to explain why the best individual is better than other alternatives.

This type of display is obviously specific to any particular problem. However, any program with an evaluation function should be able to show specified intermediate calculations in that function.

While in the edit mode I have the option to look at any single gene, or any pair of genes to see what values occur with changes over the full range of these genes (with all other genes held constant). This helps to check how much influence a given gene has on the current system, as well as checking whether it is at the true optimum. The 2-gene system is particularly useful here, but I can’t think of a good way of showing 3 or more genes at once.

The above display can either show the actual values, or use a 16 shade graph as described previously.
The 2-gene graph often shows diagonal ridges, where changing any single gene gives a worse rather than better result, whereas changing both genes can lead up the diagonal ridge to better values. I presume the same diagonal ridges occur in higher dimensions.

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.

A7: It is important not to get locked into using GAs for problems where simpler (faster) methods will do as well or better. It is also important to be satisfied that a GA is best for some problems. I have included the option to carry out several varieties of hill-climbing (and simulated annealing), and this has convinced me to stick with GAs as the main method.

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

A7: No. I would not object to being contacted in the future.
GA Visualization, Design Questionnaire.

Trevor Collins, The Knowledge Media Institute,
The Open University, Milton Keynes MK7 6AA.

Respondent - A8

*Q.1: HOW LONG HAVE YOU BEEN USING GAs?
A8: About 7 years

*Q.2: DURING THIS TIME WHAT HAVE YOU USED GAs FOR?
A8: A variety of scientific problems

*Q.3: WHY DID YOU USE GAs FOR THIS TASK?
A8: They seemed to offer the prospect of providing better results, or equivalent results in less time, than conventional techniques.

*Q.4: WHAT ENVIRONMENT(S) DO YOU USE WHEN WORKING WITH GAs? PLEASE SPECIFY EACH COMPUTING ENVIRONMENT SEPARATELY I.E. THE COMPUTER SYSTEM, PROGRAMMING LANGUAGE AND/OR APPLICATION TOOL?
A8: Generally HP 9000 workstations running HP-UX 9.01, programming in C. We do not use commercial or shareware packages, but write all our own software.

*Q.5: WHAT DO YOU FIND DIFFICULT, IF ANYTHING, ABOUT THE FOLLOWING SET-UP STEPS INVOLVED IN CREATING A GA:

*Q.5.1: DEFINING THE MAPPING BETWEEN THE PROBLEM DOMAIN AND THE STRING REPRESENTATION USE BY THE GA?
A8: This is not necessarily difficult, but clearly important. We have several times used multi-dimensional GA strings (on studies of the movement of air pollution and, more recently, studies on the analysis of liquid waste) since these provide better results for certain types of problem. We often
use non-standard coding.

*Q.5.2: PRODUCING AN EFFECTIVE EVALUATION FUNCTION?
A8: This can be a difficulty in many scientific problems; scaling is often necessary to ensure the algorithm does not concentrate on one variable and neglect others. Usually we find that in principle it is not too difficult to construct a suitable function, but often it must be refined once we know the behaviour of the algorithm.

*Q.5.3: CHOOSING THE GA's COMPONENTS, E.G. THE INITIAL POPULATION CREATION METHOD, WHAT REPRODUCTION GENE-POOL SELECTION CRITERION TO ADOPT, WHICH GENETIC OPERATORS TO APPLY, ETC.?
A8: A bit of trial and error is often required. One becomes more familiar with certain strategies, and I suppose one tends to favour those strategies, perhaps unreasonably, over others. Members of my group have a pretty free hand, and are usually eager to investigate any different approaches they can find, and not be guided much by my own experience!

*Q.5.4: SELECTING SUITABLE PARAMETERS FOR THE GA, E.G. THE POPULATION SIZE, THE MUTATION RATE (IF APPROPRIATE), ETC.?
A8: Trial and error, starting from parameters which past experience suggests will be productive.

*Q.5.5: ARE THERE ANY OTHER SET-UP STEPS THAT YOU USE BEFORE RUNNING THE GA? IF SO PLEASE NOTE THEM AND ANY ASSOCIATED DIFFICULTIES YOU ENCOUNTER BELOW.
A8:

*Q.6: HAVING APPLIED A GA TO A PARTICULAR PROBLEM WHAT APPROACH DO YOU TAKE, IN ORDER TO:

*Q.6.1: ASSESS THE QUALITY OF ANY SOLUTION(S) FOUND?
A8: Comparison with literature results if available. Comparison with results yielded by conventional approaches on the same data. Statistical analysis of the results yielded by the GA. Comparison between results of repeated runs.

*Q.6.2: EXAMINE HOW REPRESENTATIVE THE OUTPUT OF THE GA IS IN TERMS OF ALL THE POSSIBLE POINTS WITHIN THE PROBLEM-SPACE?


*Q.7: IF THE FOLLOWING TYPICAL OUTPUT CHARACTERISTICS WERE TO BE REPRESENTED WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.7.1.A: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - ADVANTAGES.

A8: None, unless the population were very small. It is often useful to have a measure of the diversity of the population, but one (or several) numerical values representing this would be preferable in most instances to viewing data on 50 or 100 individual strings.

*Q.7.1.D: ALL OF THE INDIVIDUAL CHROMOSOMES WITHIN EACH POPULATION - DISADVANTAGES.

A8: Too much screen clutter

*Q.7.2.A: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - ADVANTAGES.

A8: Better. Less screen clutter

*Q.7.2.D: A USER DEFINED SELECTION OF REPRESENTATIVE CHROMOSOMES - DISADVANTAGES.
A8: None

*Q.7.3.A: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - ADVANTAGES.
A8: A standard method of following the progress of the calculation. Generally gives a useful idea of how things are going.

*Q.7.3.D: THE RATE OF CHANGE IN THE POPULATIONS FITNESS VALUES, I.E. THE GRADIENT VALUES OF A FITNESS VERSUS GENERATION GRAPH - DISADVANTAGES.
A8: One often wants a more detailed understanding of what is happening in the population than this graph can give.

*Q.8: AS WELL AS DIRECTLY ILLUSTRATING THE OUTPUT OF THE GA, VISUALIZATION COULD BE USED TO REPRESENT ADDITIONAL INFORMATION EITHER DERIVED FROM THE OUTPUT DATASET OR RECORDED SEPARATELY. IF VISUALIZATION WERE USED TO REPRESENT THE FOLLOWING CHARACTERISTICS WHAT ADVANTAGES OR DISADVANTAGES, IF ANY, COULD YOU FORESEE?

*Q.8.1.A: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - ADVANTAGES.
A8: Depends upon the problem being tackled. We have found such a visualization useful at times.

*Q.8.1.D: THE CHROMOSOMES IN THE REPRODUCTIVE GENE-POOL - DISADVANTAGES.
A8: None

*Q.8.2.A: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION OPERATOR HAS BEEN APPLIED - ADVANTAGES.
A8: limited

*Q.8.2.D: THE OCCURRENCE OF MUTATION IN CHROMOSOMES WHERE A MUTATION
OPERATOR HAS BEEN APPLIED - DISADVANTAGES.

A8: Since mutation normally causes little change in the string, there wouldn't be a great deal to show! There should be no value in showing the position of mutation, unless for some reason one biases the position. I can't see this being very useful.


A8: limited


A8: Again this would be of interest in illustrating how the GA works, but I think of little value in helping one monitor the action of the algorithm.

*Q.8.4.A: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT’S BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - ADVANTAGES.

A8: We’ve used this type of measure a lot. Useful.

*Q.8.4.D: A “SIMILARITY” RATING FOR EACH CHROMOSOME BASED ON HOW LITTLE THEY DIFFERED TO THE FITTEST CHROMOSOME, E.G. A TEN BIT BINARY CHROMOSOME THAT DIFFERED FROM THE FITTEST CHROMOSOME IN THREE IF IT’S BIT POSITIONS (“LOCI”) MAY HAVE A SIMILARITY RATING OF 0.7 - DISADVANTAGES.

A8: None

*Q.9: PLEASE SPECIFY ANY OTHER DIRECT OR INDIRECT CHARACTERISTICS THAT
YOU WOULD BE INTERESTED IN SEEING VISUALIZED.

A8: Varies greatly from one application to the next. The most useful factors we follow relate to the degree of diversity within the population.

*Q.10: HOW HELPFUL, OR DESTRUCTIVE, WOULD YOU FIND THE FOLLOWING INTERACTION OPPORTUNITIES FOR YOUR USE OF GAs?

*Q.10.1: EXECUTION CONTROL THROUGH THE USE OF A CONTROL PANEL TO RUN, PAUSE STEP FORWARD, STEP BACKWARD, SAVE A SNAPSHOT, AND/OR STOP EXECUTION:

A8: useful

*Q.10.2: EDITING THE ALGORITHM'S PARAMETERS DURING EXECUTION:

A8: of some interest

*Q.10.3: EDITING THE POPULATION'S CHROMOSOMES BETWEEN TWO GENERATIONS:

A8: of minor value

*Q.10.4: EDITING THE REPRODUCTION GENE-POOL'S CHROMOSOMES WITHIN A GENERATION:

A8: of minor value

*Q.11: PLEASE SPECIFY ANY OTHER FORMS OF INTERACTION THAT YOU WOULD CONSIDER BENEFICIAL.

A8:

*Q.12: DO YOU HAVE ANY OTHER SUGGESTIONS ON HOW GAs COULD BE MADE EASIER TO USE? OR ANY OTHER COMMENTS AT ALL ABOUT GAs? PLEASE NOTE THEM BELOW.
A8:

*Q.13: FINALLY, WOULD YOU HAVE ANY OBJECTION TO BEING CONTACTED IN THE FUTURE WITH REFERENCE TO THIS PROJECT AND THE EVALUATION OF THE RESULTING GA VISUALIZATION SYSTEMS?

A8: No. I would not object to being contacted in the future.