

Diagnostic assessments of student thinking about stochastic processes.

Michael W. Klymkowsky^{1*}, Katja Koehler² & Melanie M. Cooper³

¹ Molecular, Cellular & Developmental Biology, University of Colorado Boulder, Boulder, Colorado 80309-0347 U.S.A.

² Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH), Zürich, Switzerland

³ Chemistry, Michigan State University, East Lansing, Michigan, 48824 U.S.A.

Abstract:

A number of research studies indicate that students often have difficulties in understanding the presence and/or the implications of stochastic processes within biological systems. While critical to a wide range of phenomena, the presence and implications of stochastic processes are rarely explicitly considered in the course of formal instruction. To help instructors identify gaps in student understanding, we have designed and tested six open source activities covering a range of scenarios, from death rates to noise in gene expression, that can be employed, alone or in combination, as diagnostics to reveal student thinking as a prelude to the presentation of stochastic processes within a course or a curriculum.

Introduction:

Randomness can be considered from a number of distinct perspectives (Eagle, 2005). A truly random event would be one that has no cause and so would be completely unpredictable. Such events would be expected to display no regularity what so ever, no matter how many similar events were considered, assuming that similar events occur. Such events can be referred to as “acts of god” (miracles) and cannot, by their very nature, be studied scientifically, a point made explicitly in Thomas Paine’s (1794) *The Age of Reason*. In contrast, there are processes that, while unpredictable at the level of individual events are well behaved (predictable) at the level of larger populations. The presumption is that such processes, commonly referred to as stochastic, are due to as yet unknown, unknowable, or uncalculatable causes. While individual events are independent of one another, individual measurements can be added together in order to arrive at a predictive model for the behavior of a larger population, a behavior known as the “law of large numbers” (Metz, 1998; Tversky and Kahneman, 1971).

There are two generic types of predictably random processes. The first are based on the effects of external factors while the second are due to internal factors. Brownian motion is an example of an externally driven stochastic

process. The erratic movements of a molecule or particle at the microscopic level are driven by theoretically deterministic but pragmatically unknowable molecular level collisions.¹ Our ability to predict the outcome of these molecular level processes is limited by the large number of individual events involved and, in some cases, theoretical limitations such as the Heisenberg Uncertainty principle. While the exact path of a particle exhibiting Brownian motion is unpredictable, the bulk movement (net flux) of large numbers of such particles (diffusion) can be predicted accurately using Fick’s Law (Berg, 1993; Ozawa, 2003).² This seeming contradiction was resolved by Boltzmann and involves the consideration of entropic factors and probability (see Lebowitz, 1993). It was this insight that led Einstein to use an analysis of Brownian motion to establish the existence of atoms (and molecules) (Einstein, 1905; Einstein and Infeld, 1938).

The second general class of stochastic events involves internal factors. This is often presented in terms of a “drunken walk” or hidden variable perspective. The stochastic nature of the process is presumed to be due to factors within the object (the walker). The radioactive decay of unstable isotopes is an example of this type of process (Kossert and Nähle, 2014)³, a phenomena apparently not well understood by students (see Prather, 2005).

¹ https://en.wikipedia.org/wiki/Brownian_motion

² https://en.wikipedia.org/wiki/Fick%27s_laws_of_diffusion

³ <http://phys.org/news/2014-10-textbook-knowledge-reconfirmed-radioactive-substances.html>

In the case of such stochastic behaviors a key concept is an event's probability. We can predict the behavior of populations based on probabilistic models without a detailed understanding of the underlying processes responsible for their behavior. Much of the past research on peoples' development of an awareness of stochastic processes has focused on when the ability to distinguish deterministic from probabilistic behaviors initially appears (Piaget, 1974; Piaget and Inhelder, 1975). In fact, when, or rather whether, people come to develop an accurate understanding of stochastic processes, so that they can apply such an understanding appropriately to specific scientific scenarios or personal experiences, remains unclear (Metz, 1998).

An obvious question then is, why does an understanding of stochastic processes matter? In biological systems the generation of mutations is a ubiquitous stochastic process involved in evolutionary, developmental, and disease processes (see Baillie et al., 2011; Ionita-Laza et al., 2009). New molecular level techniques have made it possible to move beyond bulk measurements to examine and appreciate the functional significance of the various stochastic behaviors displayed by subcellular systems (genes), individual cells, cellular networks, and social systems (Weber and Buceta, 2013). Such stochastic processes influence gene expression (e.g. Elowitz et al., 2002; Vilar et al., 2003) through their effects on the initiation of transcription and translation, leading to processes such as transcriptional and translational bursting (Eldar and Elowitz, 2010; Jia and Kulkarni, 2011; Pedraza and Paulsson, 2008; Suter et al., 2011), DNA repair (Uphoff et al., 2016), the assembly and disassembly of macromolecules (which together influence molecular half-life), molecular movements, and reaction kinetics. At the cellular and organismic levels, stochastic processes underlie social behaviors such as quorum sensing, altruistic programmed cell death (Engelberg-Kulka et al., 2006; Yarmolinsky, 1995), and the emergence of other phenotypic differences in organisms with identical genotypes, including the appearance of slowly growing drug resistant "persisters", multicellularity, and the differentiation of stalk and spore cells in *Dictyostelium* (Elowitz et al., 2002; Engelberg-Kulka et al., 2006; Huettenbrenner et al., 2003;

Strassmann et al., 2000; Travisano and Velicer, 2004; Uphoff et al., 2016; You et al., 2004).

There are many implications associated with a failure to appreciate the role of stochastic processes in biological systems. Resistance to the role of random mutation as the basis of inheritable phenotypic variation, the ultimate stochastic event in biology, was a major barrier to the acceptance of evolutionary theory; models favoring various deterministic internal or external drivers, including Lamarckian processes, orthogenesis, and/or divine intervention were originally preferred as being more plausible (Bowler, 1992; Bowler, 2005). As an aside, it is in the context of the creation and flow of information that the Central Dogma, as elucidated by Crick (1970) is important; information, generated through mutation and selection, flows out from DNA rather than entering DNA from the environment or some other source (Bowler, 1992; Koonin, 2015).⁴ As noted by others, students often do not explicitly grasp this fact (see Speth et al., 2014 and references therein); similarly students do not appear to grasp the role of genetic drift and associated processes in allele loss, and conventional instruction has been reported to increase their confusion (Andrews et al., 2012). The failure to explicitly consider the origins of biological information is likely to impact student (and teacher) acceptance and understanding of naturalistic (non-theistic) evolutionary processes (Moore, 2008).⁵

There are a few obvious reasons for the difficulty in accepting a role for stochastic processes, most importantly people do not easily accept the role of random or stochastic effects in their own day-to-day lives, rather they prefer active drivers such as fate or luck, or supernatural interventions (Taleb, 2005). A classic example involves what is known as the Gambler's Fallacy (Turner, 2000; Tversky and Kahneman, 1974), a mistaken assumption that independent events are in fact dependent on one another.

A second issue is that all too often students are not explicitly introduced to stochastic processes during the course of instruction. Andrews et al (2012) note that students are often taught about genetic drift in the context of scenarios (Hardy-Weinberg equilibria) in which drift is theoretically impossible. Similarly, while the specificity and stability of a molecular interaction

⁴ While not surprising given Griffiths' discovery of bacterial transformation, the discovery of horizontal gene transfer complicates this view.

⁵ <http://www.gallup.com/poll/170822/believe-creationist-view-human-origins.aspx>

is based on energy released upon complex formation, the mechanisms behind molecular dissociation are rarely explicitly discussed in the molecular biology textbooks we have examined. Similarly, it is commonplace to find that when molecular level processes are illustrated, particularly in video animations, their stochastic nature is rarely explicitly depicted or acknowledged.⁶ Based on student responses to

questions on the Biology Concepts Instrument (BCI) (Klymkowsky et al., 2010) as well as a number of other studies, it is clear that students routinely fail to recognize the role of stochastic processes in biological systems (FIG. 1) (Andrews et al., 2012; Garvin-Doxas and Klymkowsky, 2008; Odom, 1993; Speth et al., 2009; Speth et al., 2014) (Champagne-Queloz, et al, in preparation). For

example, the ratio of unproductive to productive collisions within a cell (or in an *in vitro* system) between nucleotide triphosphate molecules and the DNA replication machinery (or the RNA synthesis complex), or between amino acid charged tRNAs and the mRNA/ribosome complex (during polypeptide synthesis) is generally presented as zero (that is, no unproductive collisions are illustrated), while the actual ratio is typically greater than 4 to 8 for DNA replication (depending on whether one considers deoxy- and ribo-nucleotides) and greater than 20 for polypeptide synthesis (a number complicated by the fact that different amino-acid-tRNAs are present in different concentrations and different codons can be used to direct the addition of the same amino acid).

In the same vein in the context of evolutionary processes, it is uncommon to see any consideration given to the effects of population size, whether from founder effects or population bottlenecks, together with subsequent genetic drift, on the presence of specific traits or a population's evolutionary trajectory (see Blount et al., 2008; Keinan et al., 2007; Lynch and Conery, 2003). Similarly, it is our impression that few courses consider the stochastic behaviors that influence gene expression and their effects on

cellular behavior (see Cottrell et al., 2012; Dar et al., 2015; Eldar and Elowitz, 2010; Jia and Kulkarni, 2011; Raj and van Oudenaarden, 2008; Swain et al., 2002; Uphoff et al., 2016; Vilar et al., 2003). This includes the recent discourse on the role of random events (luck) (Tomasetti and Vogelstein, 2015) versus environmental drivers in the origins of human cancers (Wu et al., 2015).

Once two molecules bind to one another, how could they come back apart again?

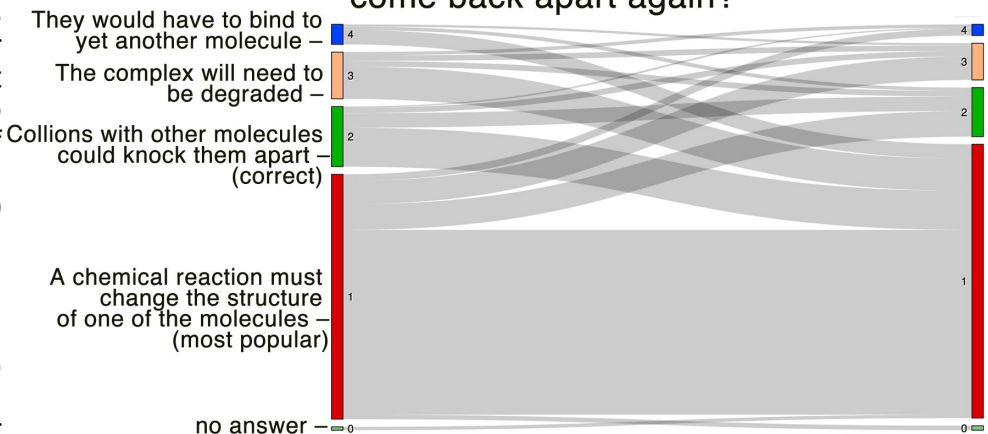


FIG. 1: In this cohort of students, there is little change (over the course of instruction) in the recognition that the dissociation of a molecular complex is driven by collisions with surrounding molecules (from Champagne-Queloz, A. Ph.D. et al., in preparation).

As part of a general effort to consider course and curricular content in chemistry and biology (see Cooper and Klymkowsky, 2013a; Cooper and Klymkowsky, 2013b; Klymkowsky and Cooper, 2012; Klymkowsky et al., 2010) and informed by student responses to various BCI questions (Garvin-Doxas and Klymkowsky, 2008; Henson et al., 2012; Klymkowsky, 2007)(unpubl. obs.)(FIG. 1) we have developed tools to gauge students' understanding of a range of stochastic processes. We use student-drawn graphs and models (Bryfczynski et al., 2012; Cooper et al., 2014; Trujillo et al., 2012; Williams et al., 2015) to extend insights gathered through multiple choice and open response questions with the goal of providing instructors and course designers with a clearer picture of student thinking in this area. This effort also reveals aspects of students' ability to convey their thoughts and construct arguments, as well as their ability to generate and interpret graphs (general numerical and analytic literacy). Two of these activities (death and

⁶ see HHMI: <http://www.hhmi.org/biointeractive/translation-basic-detail>, <http://www.hhmi.org/biointeractive/dna-transcription-advanced-detail>, DNA Learning Center: https://youtu.be/Tfyf_rPWUdY, Harvard Biovisions: <https://youtu.be/wJyUtbn0O5Y>

radioactive decay) involve intrinsic sources of stochastic behavior and four (motion, molecular dissociation, genetic drift, and noise in gene expression) reflect external drivers. Designed as diagnostics, these open source activities can be freely adapted to serve as formative assessments. For each activity, we describe what correct answers should contain and illustrate common student responses we have observed.

Methods: These activities were developed and tested using the beSocratic system running out of Michigan State University. To gauge student responses, the activities were completed by students in a revised introductory biology course at UC Boulder that emphasizes both evolutionary processes and the qualitative analysis of

background, that is, what courses they have taken previously. The activities are typically between 6 to 8 panels in length; students are supplied with necessary background information, multiple choice questions to be answered, and tasks to be completed. Students are generally asked to explain their decisions using open text responses.

Activity #1 – Predicting death (6 panels): There are a number of reasons why the ability to predict the probability of death in a population is useful, most notably in the context of insurance. Such institutions accumulate money from their subscribers and, after a certain age, pay out to their members a set stipend. Being able to accurately estimate the average life expectancy of individuals within a population is

necessary for the financial viability of such institutions. For a large enough population life expectancy data has the form of a smooth (well-behaved) curve, that is the percentage of the original population that remains alive (or has died) as a function of time is predictable. In contrast, when we consider the fate of smaller groups of people (or of a particular person), the curve has the form of a step function subject to stochastic fluctuations; in the abstract, when a particular person dies is a stochastic event. If students have not previously been introduced

to considering the probabilities of events, they will require explicit instruction on probability and its relevance in particular situations. The details of this activity are presented in supplemental data 1.1.

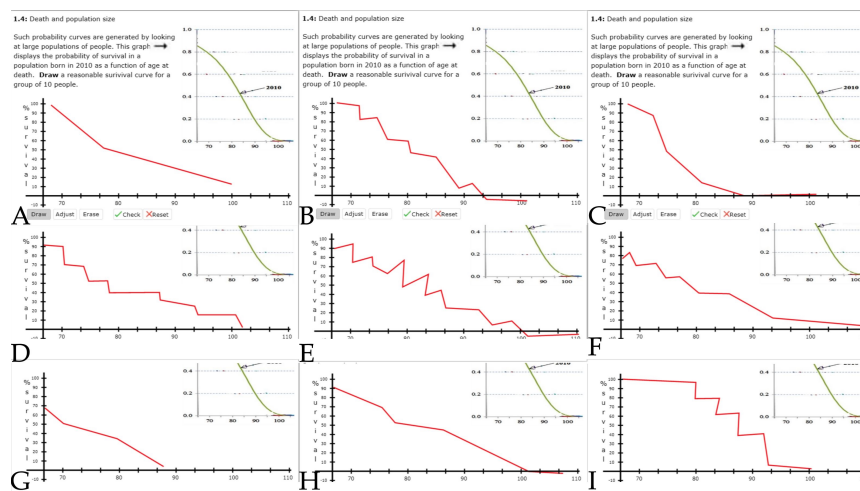


FIG. 2: Student drawings of the survival behavior of a small group of 10 people. Often only a few (as illustrated in parts B, D, E, F, and I) recognize that the graph is a step function and that different groups will generate different graphs.

biological systems; volunteers were awarded 5 extra credit points (added to their final course grade, based on 500 possible points) without regard to whether their responses were correct or not, although completion of all activities was considered in assigning points. Course grades were not curved, so student participation did not negatively impact students who chose not to participate. This activity was judged exempt (IRB protocols 0304.09 and 15-0347). Students' responses led to revisions of the activities, primarily the deletion of some pages and the consolidation of questions asked.

Description of activities and student responses: All of the activities begin with a question or two about the student's educational

Outcomes: Students are expected to be able to recognize, and predict, the differences between the behaviors of large populations (regular and predictable) and small populations (subject to stochastic variation), as witness by their drawing of step graphs which differ between different small groups (FIG. 2). Based on the shapes of survival curves as a function of time, students should be able to deduce that survival rates are influenced by changing environmental factors while the maximum lifespan is not, suggesting that maximum longevity is constrained by inherent (genetic) factors. Students are expected to consider and accurately plot data (something that, perhaps surprisingly, students have great difficulty with), and to recognize how observations of small groups can, by adding them together, be

used to make accurate predictions about the behavior of larger groups.

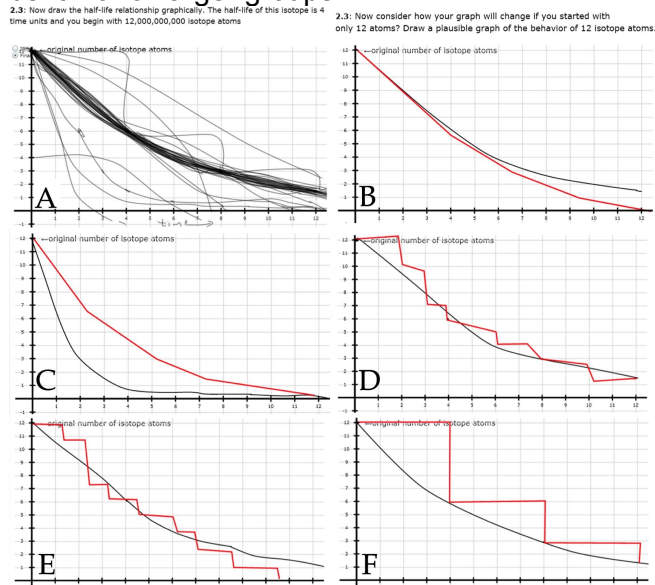


FIG. 3: Student drawings of the radioactive decay of a large group of atoms (A: all student decay graphs presented together: parts B-F graphs of individual students – large group decay behavior (in black lines) compared to the behavior of a small group of 20 atoms (red lines in parts B-F). Typically only a few students (illustrated in parts D, E, & F) recognized that the graph of small population behavior is a step function, and that different small groups of isotope atoms will generate different graphs.

Activity #2 – Predicting radioactive decay (6 panels): The radioactive decay of isotopes is perhaps the classic example of a stochastic process that can be considered at the level of individual atoms or large populations of atoms. It is likely that students have been introduced to the process of radioactive decay in the course of their education, which provides the instructor (and student) an opportunity to monitor the effectiveness of these educational experiences. If students have not previously encountered isotopes, isotope decay, or the concept of half-life their instructor will need to decide what level of background instruction is appropriate and realistic. We note that the concept of half-life (which applies to a wide-range of biomolecules) is generally not well understood by students (see Prather, 2005). The details of this activity are presented in supplemental data 1.2.

Outcomes: Students are expected to be able to predict the differences in the behavior of large and small populations of unstable isotope atoms (FIG. 3). Students are also expected to be able to predict how best to use studies of small populations to predict the behavior of larger

groups. This will almost certainly require the presentation of an event's probability as a function of time.

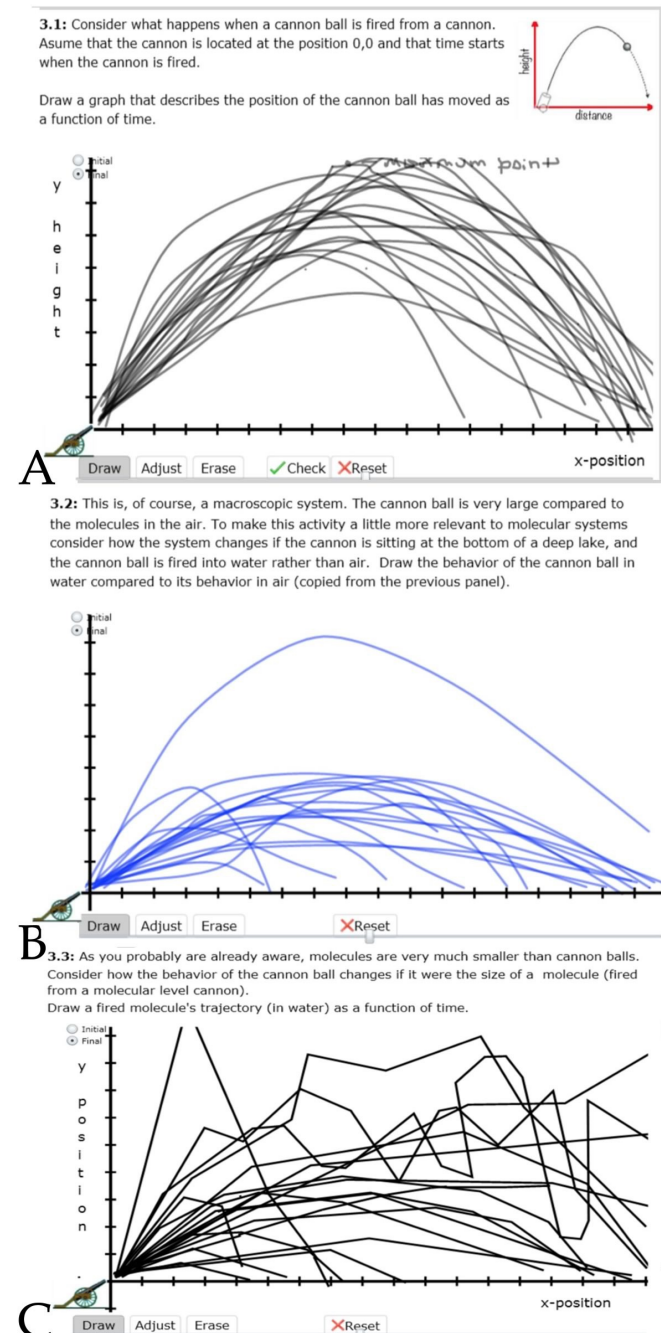


FIG. 4: Student drawings of projectile motion in air (A), water (B), and as the size of the projectile approaches the size of a molecule (C). At the molecular level (C) few students recognize the Brownian behavior of molecules.

Activity #3 – Predicting molecular trajectories (6 panels): In the context of biological systems, the kinetic movements of molecules supply the energy needed for reactions to proceed (by overcoming the activation energy), for larger

molecular complexes to move (diffuse) within the cell, and for complexes, once formed, to come apart again, a critical feature of dynamic systems. However, in most physics courses students are introduced to macroscopic behaviors, such as the movement of objects, based on the concepts of force and momentum. At the molecular level, the movement of particles will be affected by interactions and collisions with other molecules. Indeed, Einstein used a consideration of Brownian (microscopic) motion as evidence for the existence of molecules (Einstein, 1905; Einstein and Infeld, 1938). This formative activity was designed to reveal evidence about whether students can transfer their understanding of motion and interactions from the macroscopic to the molecular level. If students have not been previously introduced to the intermolecular interactions and collisions and the phenomena of Brownian motion, they will likely require instruction on these topics. However, since most students have been exposed to these topics, this activity will provide feedback on whether they are able to apply these ideas. The details of this activity are presented in supplemental data 1.3.

Outcomes: Students are expected to be able to generate plausible projectile trajectories in different contexts (e.g. air versus water) and then predict how projectile size (macroscopic versus molecular) influences movement (FIG. 4). This includes recognizing the transition to stochastic (Brownian) behavior as projectile size approaches the molecular. Students are also asked to consider how they might combine observations of small groups to make predictions about the behavior of larger groups.

Activity #4 – Predicting molecular association and dissociation (8 panels): Interactions between molecules underlie the binding interactions involved in a range of biological activities. The forces associated with electrostatic attractions are a consequence of the transient (London Dispersion Forces) and permanent dipoles associated with atoms and bonds. The strength of these intermolecular interactions is influenced both by the size and the shapes of the interacting molecules, and the interactions are overcome by the kinetic energy delivered through collisions with other molecules. Responses to this activity provide insights into students' understanding of the factors that influence molecular association, binding specificity, and dissociation rates. Intermolecular interactions are often considered in the context of introductory (general) chemistry courses, since they explain

physical properties such as melting and boiling points and solvent solubility; they are, however, often not well mastered by students (see Williams et al., 2015). If these ideas have not been well understood, their introduction or review may be necessary. The details of this activity are presented in supplemental data 1.4.

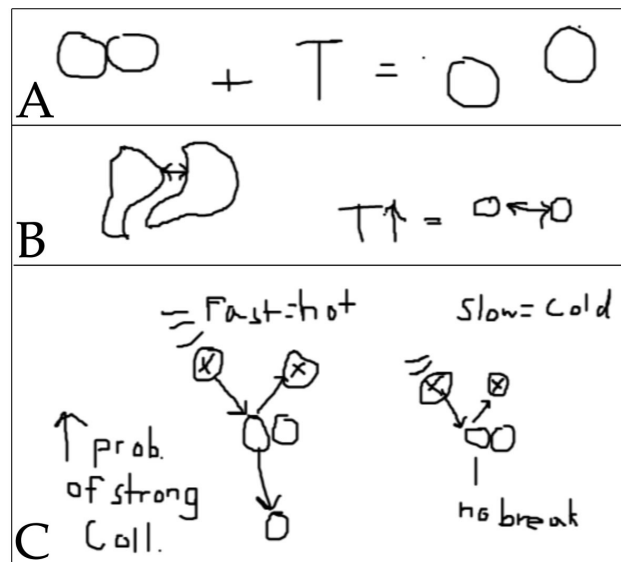


FIG.5: Students' drawings of molecular dissociation can represent temperature abstractly (A,B) or more directly (C) in the context of molecular collisions.

Outcomes: Students are expected to be able to predict (and explain the logic behind their decisions) the relative strength of interactions between molecules, based on their shapes. They are expected to recognize i) how energy is redistributed when molecules interact and to explain their logic in terms of binding (potential) energy, kinetic energy (delivered through collisions with surrounding molecules), and energy conservation; ii) the relationship between kinetic energy, molecular mass and velocity, iii) the distribution of energy in a population of molecules as a function of temperature, and iv) how changes in temperature (typically) influences the stability (dissociation) of a molecular complex (FIG. 5). Entropic effects, associated with the $-T\Delta S$ term in the free energy equation, are ignored here.

Activity #5 – Genetic drift (8 panels): Alleles are generated by the stochastic process of mutation (as opposed to various forms of site directed mutagenesis). Within a population the relative frequency of the alleles at a particular genetic locus is the result of various adaptive (natural, sexual, and social selection) and non-adaptive (genetic drift, founder effects, population

bottlenecks, and gene linkage) effects. In the case of genetic drift, the allele frequencies within a population are influenced by stochastic processes. In sexually reproducing species, these processes include which alleles are delivered to which gametes and which gametes combine to form the next generation.

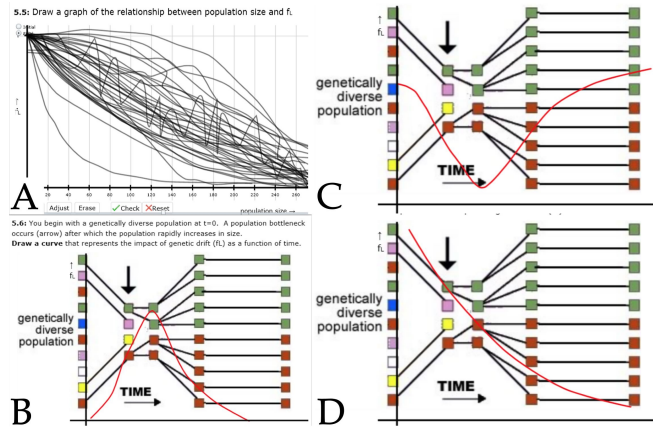


FIG. 6: **A:** The ensemble of student graphs (black lines) of F_L as a function of population size. Individual student graphs (red lines) of how F_L changes as a function of a change (a dramatic decrease) in population size. **B** reflects an accurate graph, **C** indicates a change in F_L in the wrong direction, while **D** suggests an alternative (ambiguous) conception.

Our consideration of genetic drift is based on a Java-based web application that can be used by students to generate their own data sets; it does not explicitly consider the processes involved in producing genetic drift.⁷ Because running Java can pose security complications on some computers, we supply students with a data set (obtained from this applet) in the context of the activity. The details of this activity are presented in supplemental data 1.5.

Outcomes: The genetic drift activity expects students to be able to describe their understanding of genetic drift and to recognize when an allele has been lost or fixed within a population. Students are expected to be able to calculate the observed frequency of allele loss (F_L) as a function of population size, to plot that data, and to explain why it makes sense to draw a “best fit” curve rather than connecting the empirically derived values of F_L . They are expected to be able to predict (based on the relationship between F_L and population size) the effects of drastic changes in population size on

the value of F_L as well as to speculate on the impact of genetic drift on future evolution and to explain their reasoning. In this context, it is worth noting that Andrews et al (2012) reported that distinct misconceptions related to genetic drift not only persisted but increased in frequency after instruction on genetic drift. This provides a rationale for examining student thinking using this or another activity (see Price et al., 2014) in a pre-/post-instruction assessment model.

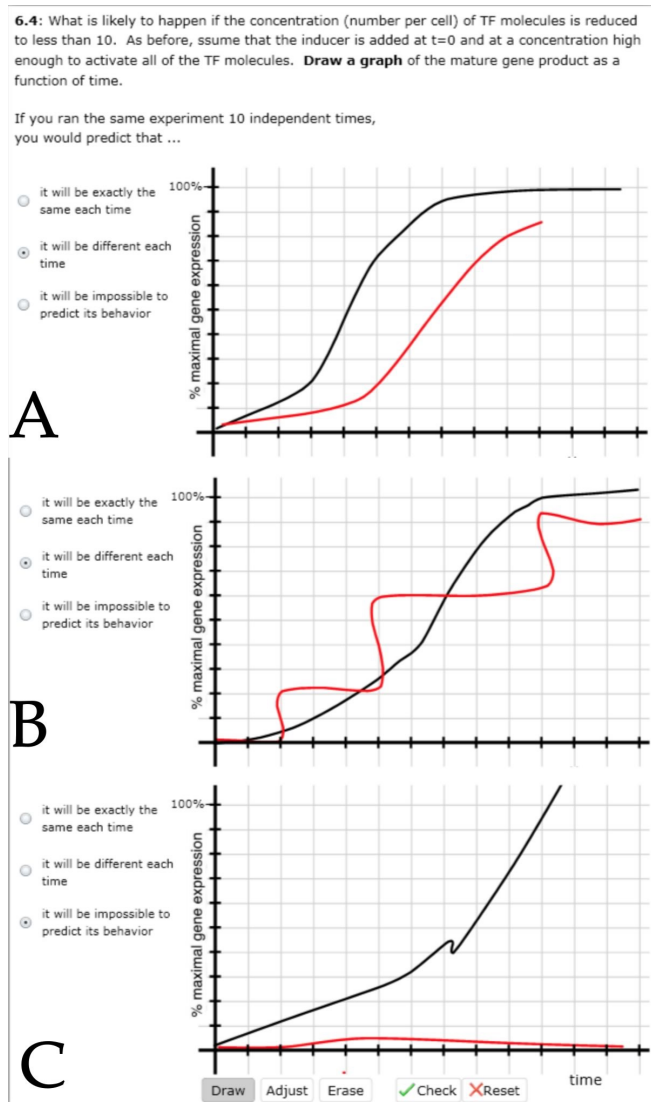


FIG. 7: Students are asked to compare (graphically) the effects of high (black) and low (red) concentrations of a transcription factor. **A** depicts similar patterns for both, **B** depicts a more stochastic pattern of expression, while **C** suggests that there is no transcription at the low transcription factor concentration.

⁷ see <http://darwin.eeb.uconn.edu/simulations/jdk1.0/drift.html>

Activity #6 – Noisy gene expression (5 panels):

This activity considers the various stochastic factors that influence gene expression. Noisy gene expression, a process that includes transcription, translation, assembly, and localization, is a phenomenon that has at its root in the fact that most cells have only two copies of any particular gene and that the number of molecules regulating the expression of a particular gene, while varying widely, remains relatively small, ranging from less than 10 per cell (in the case of the lac repressor) to many thousands.⁸ Outcomes from this activity can reveal how students approach questions related to individual gene and cell behavior. It is assumed that students already have an understanding of the basics of gene expression, including the processes that influence binding and dissociation (Activity 4) of transcription factors with DNA, as well as the interactions between the mRNA polymerase and DNA (responsible for the termination of transcription), mRNA and splicing factors and the nuclear pore/cytoplasmic transport complex, and the ribosomal complex, including the role of stop codons and release factors in the termination of translation. The details of this activity are presented in supplemental data 1.6.

Outcomes: Students are expected to be able to describe (and diagram) the processes associated with gene expression (the binding of transcription factors, the recruitment and activation of RNA polymerase, the opening of the DNA double helix, etc.). They are expected to be able to generate a plausible graphical representation of the appearance of a gene product in a cell that contains a large (10,000) number of the positively acting transcription factors that regulate the target gene, and then a comparable representation of how reducing the number of those transcription factors to 10 will influence their graph. Based on their responses, students rarely explicitly indicate the times required for transcription and mRNA processing to occur. While translation can begin before transcription is complete in prokaryotes, eukaryotic genes often have many large introns that must be removed (spliced out) before the mRNA is transported into the cytoplasm and can interact with the translational machinery. Finally students are asked to predict and justify how the numbers of transcription factor molecules will influence transcriptional noise.

Here are two examples of correct responses to the effect of changing transcription factor concentrations on transcription.

“The more transcription factors are present, the more likely they will bind to the DNA and begin the process. If there are not as many, the probability of binding will be much lower and the process will be noisier.”

“With higher concentrations of TF, the probability of it colliding with the correct sequence increases and therefore decreases the random noise”.

Examples of incorrect responses: *“depends on rna polymerase because that is what creates the gene’s expression. Transcription factors are only in transcription and the beginning of translation”*

“Higher concentration means that more molecules are colliding and moving around the cell, making more noise.”

Discussion:

The six diagnostic activities described here can serve as a means to monitor students’ thinking about a range of stochastic processes. It is not our intention that any particular course would use all six activities, since different courses focus on different content areas. That said, all of these processes are deeply analogous, with similar conceptual foundations; they offer an opportunity to determine whether students’ understanding of stochastic processes in one context is transferred to other contexts during a course or within a curriculum.

Based on students’ responses to one or more of these activities, changes in instructional design, both in terms of the content presented and stressed and skills practiced can be considered and their effects monitored. In our own case, it is clear that while the course within which these activities were initially tested sought to improve student understanding of the presence and implications of stochastic processes, there are areas in which further instructional redesign, particularly student time on task and constructive feedback, are needed.

While designed as diagnostics, the activities presented here can be used as teaching tools. For example, while stochastic processes are the common theme, different activities include

⁸ see the B10NUMB3R5 site: <http://bionumbers.hms.harvard.edu>

other important concepts, such as the difference between environmental factors and evolved traits (activity #1), radioactive decay (activity #2), the factors that control molecular association and dissociation (activity #3), molecular movement (activity #4), non-adaptive evolutionary processes (activity #5), and gene regulatory processes (activity #6). A course designer or instructor could adapt these activities to fit the specific context of their course or curriculum. What is critical is that they can reveal lacunae in students' understanding or ability to apply stochastic processes. While the analysis of textual responses and drawings is, by its very nature, more time consuming, the results of the associated multiple choice questions and student generated graphs can be quickly evaluated and even used as exemplars within a class, while the more time-consuming analyses of students' textual responses can be used as part of a redesign process, refocusing course goals, materials, and activities. A tangential benefit associated with having students generate and interpret graphs, as well as open text responses is that it can reveal the presence of previously

unexpected issues with students' ability to analyze data, generate and interpret graphs, or to construct logical, empirically-based arguments and conceptual justifications.

Ultimately, the use of these and similar activities can be expected to reinforce the importance of stochastic processes in biological systems, and to help build courses that acknowledge their importance and the conceptual difficulties associated with grasping them in an accurate and useable manner. Ignoring these aspects of biological systems, which has been all too common, encourages a rigid deterministic view of living systems that is directly contradicted by how these systems actually behave and evolve.

Acknowledgements: *This work was supported in part through NSF grants DUE-1043707 and 1341987 (M.M. Cooper, PI, M.W. Klymkowsky, Co-PI). We acknowledge various ideas supplied through conversations with Natalie Jones (UC Boulder) who helped build a genetic drift activity for the Biofundamentals course.*

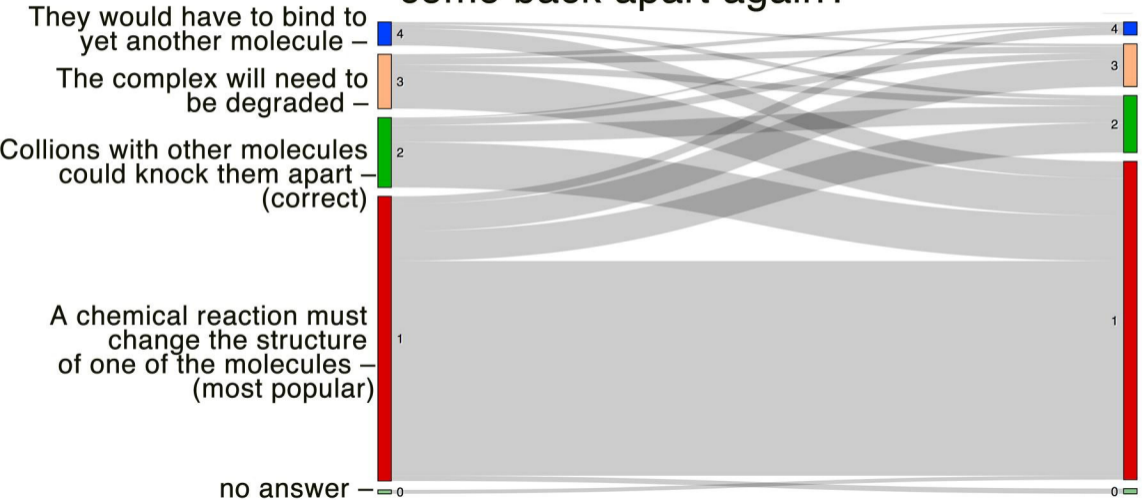
Literature cited

- Andrews, T., Price, R., Mead, L., McElhinny, T., Thanukos, A., Perez, K., Herreid, C., Terry, D. and Lemons, P. (2012). Biology undergraduates' misconceptions about genetic drift. *CBE-Life Sciences Education* 11, 248-259.
- Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., Brennan, P. M., Rizzu, P., Smith, S. and Fell, M. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534-537.
- Berg, H. C.** (1993). *Random walks in biology*: Princeton University Press.
- Blount, Z. D., Borland, C. Z. and Lenski, R. E.** (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A* 105, 7899-7906.
- Bowler, P. J.** (1992). The eclipse of Darwinism: Anti-Darwinian evolution theories in the decades around 1900. Baltimore: Johns Hopkins University Press.
- (2005). Revisiting the Eclipse of Darwinism. *J. Hist. Biol.* 38, 19-32.
- Bryfczynski, S., Pargas, R. P., Cooper, M. M. and Klymkowsky, M. W.** (2012). BeSocratic: Graphically-assessing Student Knowledge. In International Association for Development of the Information Society Conference on Mobile Learning. Berlin, Germany.
- Cooper, M. M. and Klymkowsky, M. W.** (2013a). Chemistry, life, the universe, and everything: a new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education* 90, 1116-1122.
- Cooper, M. M. and Klymkowsky, M. W.** (2013b). The trouble with chemical energy: why understanding bond energies requires an interdisciplinary systems approach. *CBE Life Sci Educ* 12, 306-312.
- Cooper, M. M., Underwood, S. M., Bryfczynski, S. P., Klymkowsky, M. W., Cole, R. and Bunce, D.** (2014). A Short History of the Use of Technology to Model and Analyze Student Data for Teaching and Research. *Tools of Chemistry Education Research* 1166.
- Cottrell, D., Swain, P. S. and Tupper, P. F.** (2012). Stochastic branching-diffusion models for gene expression. *Proceedings of the National Academy of Sciences* 109, 9699-9704.
- Crick, F.** (1970). Central dogma of molecular biology. *Nature* 227, 561-563.
- Dar, R. D., Razooky, B. S., Weinberger, L. S., Cox, C. D. and Simpson, M. L. (2015). The Low Noise Limit in Gene Expression. *PLoS one* 10, e0140969.
- Eagle, A.** (2005). Randomness is unpredictability. *The British Journal for the Philosophy of Science* 56, 749-790.
- Einstein, A.** (1905). On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der Physik*, 17, 549-560.
- Einstein, A. and Infeld, L. (1938). *The evolution of physics*. New York: Norton.

- Eldar, A. and Elowitz, M. B.** (2010). Functional roles for noise in genetic circuits. *Nature* **467**, 167-173.
- Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* **297**, 1183-1186.
- Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I. and Hazan, R.** (2006). Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genet* **2**, e135.
- Garvin-Doxas, K. and Klymkowsky, M. W.** (2008). Understanding Randomness and its impact on Student Learning: Lessons from the Biology Concept Inventory (BCI). *Life Science Education* **7**, 227-233.
- Henson, K., Cooper, M. M. and Klymkowsky, M. W.** (2012). Turning randomness into meaning at the molecular level using Muller's morphs. *Biol Open* **1**, 405-410.
- Huettenbrenner, S., Maier, S., Leisser, C., Polgar, D., Strasser, S., Grusch, M. and Krupitza, G. (2003). The evolution of cell death programs as prerequisites of multicellularity. *Mutat Res* **543**, 235-249.
- Ionita-Laza, I., Rogers, A. J., Lange, C., Raby, B. A. and Lee, C.** (2009). Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* **93**, 22-26.
- Jia, T. and Kulkarni, R. V.** (2011). Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. *Physical review letters* **106**, 058102.
- Keinan, A., Mullikin, J. C., Patterson, N. and Reich, D.** (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature genetics* **39**, 1251-1255.
- Klymkowsky, M. W.** (2007). Conceptual interference in biology education: How jigsaw puzzle/lock and key models of molecular interactions impact understanding evolutionary change. In *National Association for Research in Science Education*, pp. P-269-1554-1553-1583. New Orleans, LA.
- Klymkowsky, M. W. and Cooper, M. M.** (2012). Now for the hard part: the path to coherent curricular design. *Biochem Mol Biol Educ* **40**, 271-272.
- Klymkowsky, M. W., Underwood, S. M. and Garvin-Doxas, K.** (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. In *arXiv*: Cornell University Library.
- Koonin, E. V.** (2015). Why the Central Dogma: on the nature of the great biological exclusion principle. *Biology direct* **10**, 1-5.
- Kossert, K. and Nöhle, O. J.** (2014). Long-term measurements of 36 Cl to investigate potential solar influence on the decay rate. *Astroparticle Physics* **55**, 33-36.
- Lebowitz, J. L.** (1993). Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy. *Physica A: Statistical Mechanics and its Applications* **194**, 1-27.
- Lynch, M. and Conery, J. S.** (2003). The origins of genome complexity. *Science* **302**, 1401-1404.
- Metz, K. E.** (1998). Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction* **16**, 285-265.
- Moore, R.** (2008). Creationism in the biology classroom: what do teachers teach & how do they teach it? *The American Biology Teacher* **70**, 79-84.
- Odom, A. L., Barrow, L. H.** (1993). Freshman biology non-major students' misconceptions about diffusion and osmosis concepts. In Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics (ed. J. Novak). Ithaca, New York: Cornell University (distributed electronically).
- Ozawa, M.** (2003). Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement. *Physical Review A* **67**, 042105.
- Paine, T.** (1794). *The Age of Reason: Being an investigation of true and fabulous theology*. London: Barriois.
- Pedraza, J. M. and Paulsson, J.** (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**, 339-343.
- Piaget, J.** (1974). *Understanding causality*. New York: W. W. Norton & Co. Inc.
- Piaget, J. and Inhelder, B.** (1975). The origin of the idea of chance in children. (Trans L. Leake, P. Burrell & HD Fishbein): WW Norton.
- Prather, E.** (2005). Students' beliefs about the role of atoms in radioactive decay and half-life. *Journal of Geoscience Education* **53**, 345.
- Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A. and Perez, K. E.** (2014). The Genetic Drift Inventory: a tool for measuring what advanced undergraduates have mastered about genetic drift. *CBE-Life Sciences Education* **13**, 65-75.
- Raj, A. and van Oudenaarden, A.** (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-226.
- Speth, E. B., Long, T. M., Pennock, R. T. and Ebert-May, D.** (2009). Using Avida-ED for teaching and learning about evolution in undergraduate introductory biology courses. *Evolution: Education and Outreach* **2**, 415-428.
- Speth, E. B., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R. and Long, T.** (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE-Life Sciences Education* **13**, 529-539.
- Strassmann, J. E., Zhu, Y. and Queller, D. C.** (2000). Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965-967.
- Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472-474.
- Swain, P. S., Elowitz, M. B. and Siggia, E. D.** (2002). Intrinsic and extrinsic contributions to stochasticity in

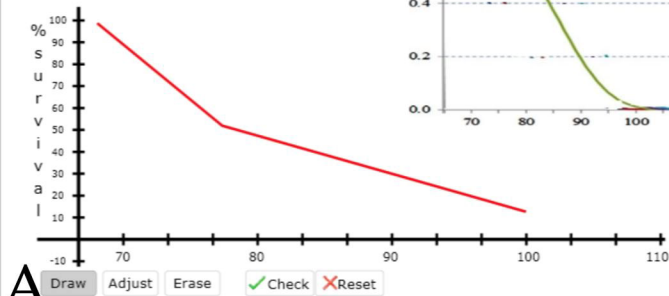
- gene expression. *Proc Natl Acad Sci U S A* **99**, 12795-12800.
- Taleb, N. N.** (2005). Fooled by Randomness: The hidden role of chance in life and in the markets. (2nd edn). New York: Random House.
- Tomasetti, C. and Vogelstein, B.** (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78-81.
- Travisano, M. and Velicer, G. J.** (2004). Strategies of microbial cheater control. *Trends in microbiology* **12**, 72-78.
- Trujillo, C., Cooper, M. M. and Klymkowsky, M. W.** (2012). Using graph-based assessments within socratic tutorials to reveal and refine students' analytical thinking about molecular networks. *Biochem Mol Biol Educ* **40**, 100-107.
- Turner, N.** (2000). Randomness, Does It Matter? *Journal of Gambling Issues*.
- Tversky, A. and Kahneman, D.** (1971). Belief in the law of small numbers. *Psychological bulletin* **76**, 105.
- (1974). Judgment under uncertainty: Heuristics and biases. *science* **185**, 1124-1131.
- Uphoff, S., Lord, N. D., Okumus, B., Potvin-Trottier, L., Sherratt, D. J. and Paulsson, J.** (2016). Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science* **351**, 1094-1097.
- Vilar, J. M., Guet, C. C. and Leibler, S.** (2003). Modeling network dynamics: the lac operon, a case study. *J Cell Biol* **161**, 471-476.
- Weber, M. and Buceta, J.** (2013). Dynamics of the quorum sensing switch: stochastic and non-stationary effects. *BMC systems biology* **7**, 6.
- Williams, L. C., Underwood, S. M., Klymkowsky, M. W. and Cooper, M. M.** (2015). Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *Journal of Chemical Education*.
- Wu, S., Powers, S., Zhu, W. and Hannun, Y. A.** (2015). Substantial contribution of extrinsic risk factors to cancer development. *Nature*.
- Yarmolinsky, M. B.** (1995). Programmed cell death in bacterial populations. *Science* **267**, 836.
- You, L., Cox, R. S., 3rd, Weiss, R. and Arnold, F. H.** (2004). Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868-871.

Once two molecule bind to one another, how could they come back apart again?

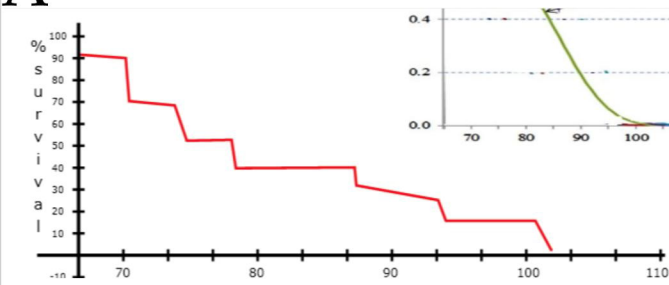


1.4: Death and population size

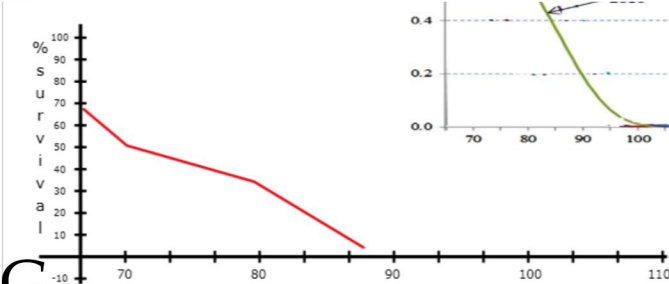
Such probability curves are generated by looking at large populations of people. This graph → displays the probability of survival in a population born in 2010 as a function of age at death. **Draw** a reasonable survival curve for a group of 10 people.



A



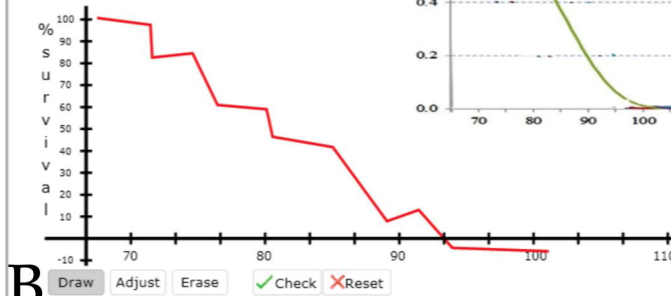
D



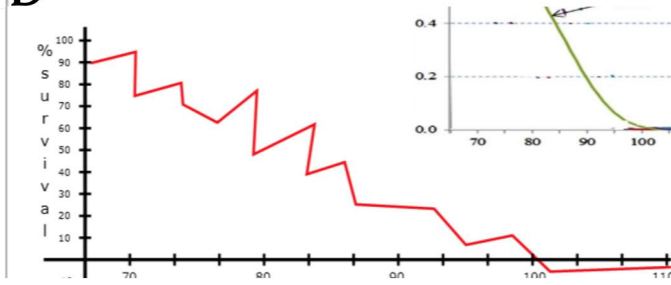
G

1.4: Death and population size

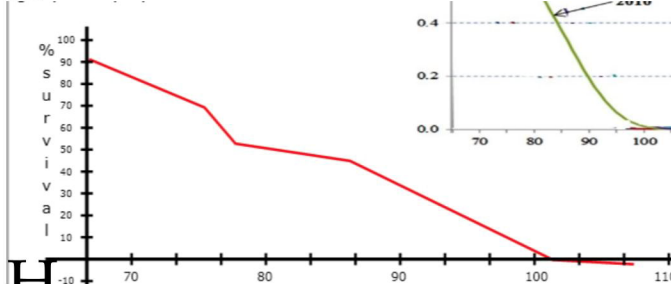
Such probability curves are generated by looking at large populations of people. This graph → displays the probability of survival in a population born in 2010 as a function of age at death. **Draw** a reasonable survival curve for a group of 10 people.



B



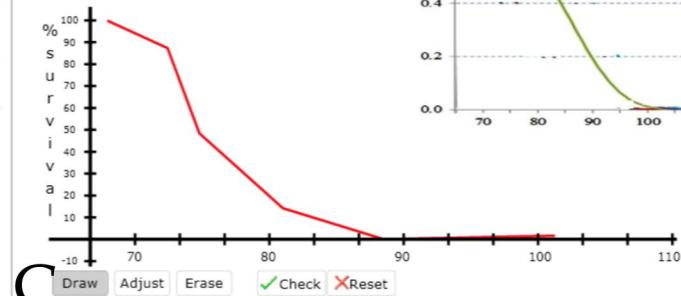
E



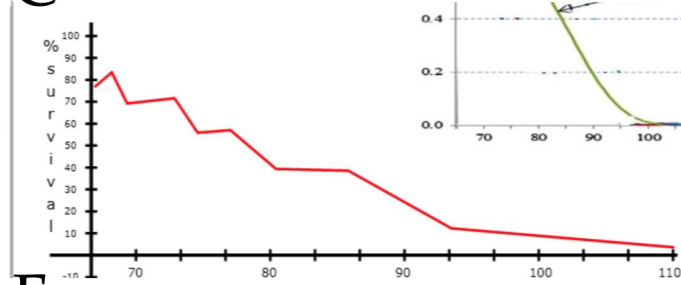
H

1.4: Death and population size

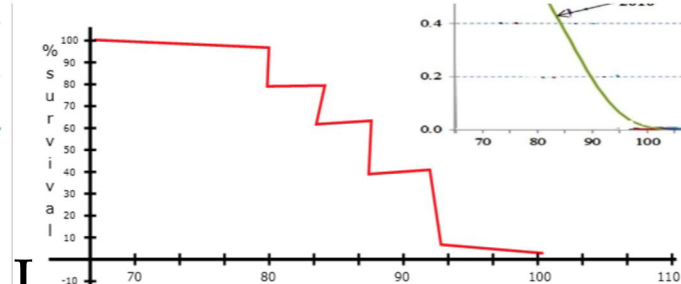
Such probability curves are generated by looking at large populations of people. This graph → displays the probability of survival in a population born in 2010 as a function of age at death. **Draw** a reasonable survival curve for a group of 10 people.



C

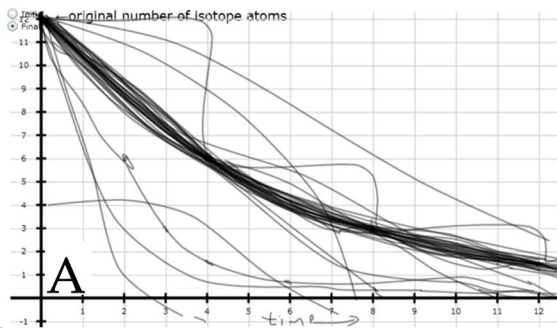


F

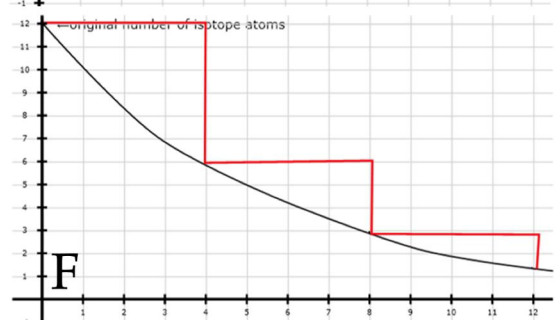
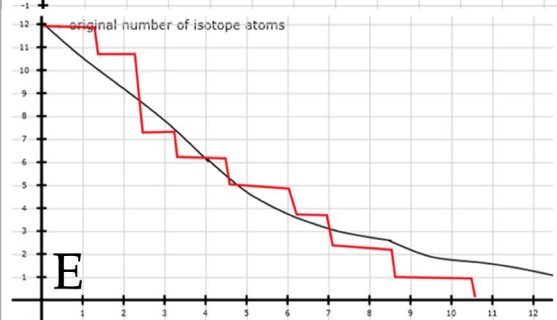
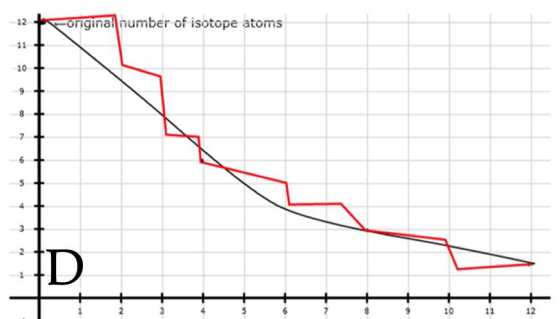
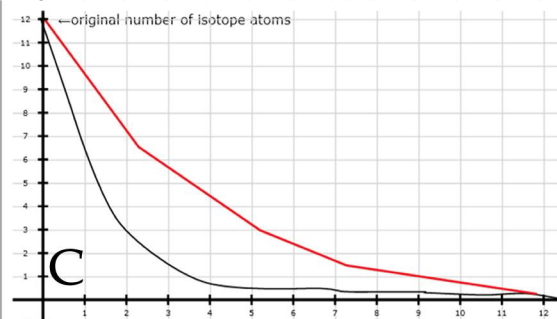
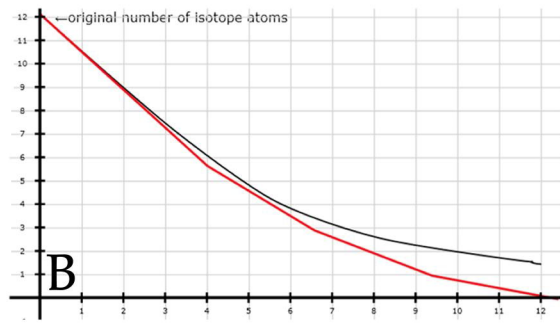


I

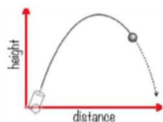
2.3: Now draw the half-life relationship graphically. The half-life of this isotope was 4 time units and you begin with 12,000,000,000 isotope atoms



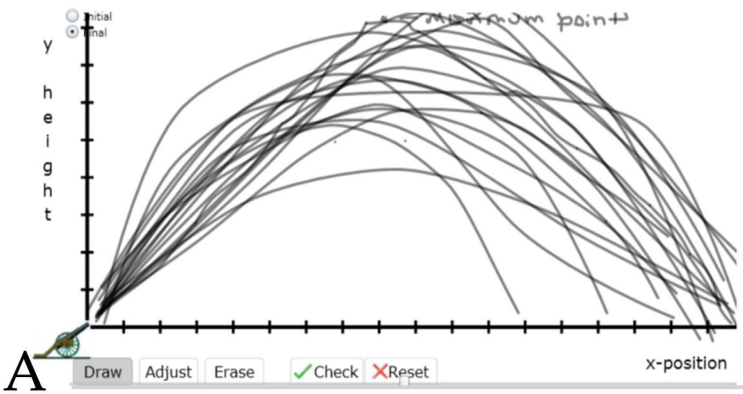
2.3: Now consider your graph will change if you started with only 12 atoms? Draw a plausible graph of the behavior of 12 isotope atoms.



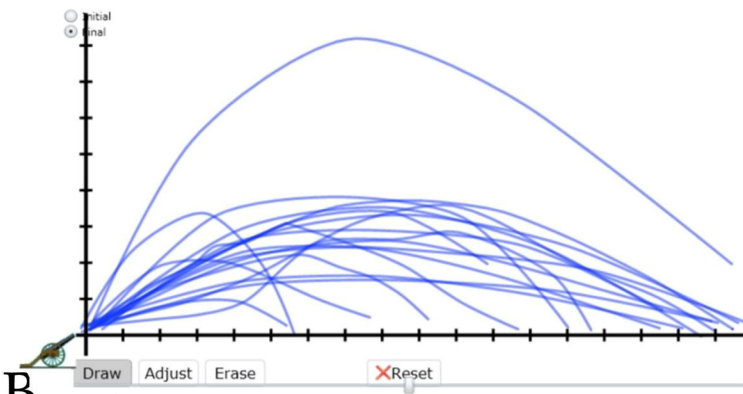
3.1: Consider what happens when a cannon ball is fired from a cannon. Assume that the cannon is located at the position 0,0 and that time starts when the cannon is fired.



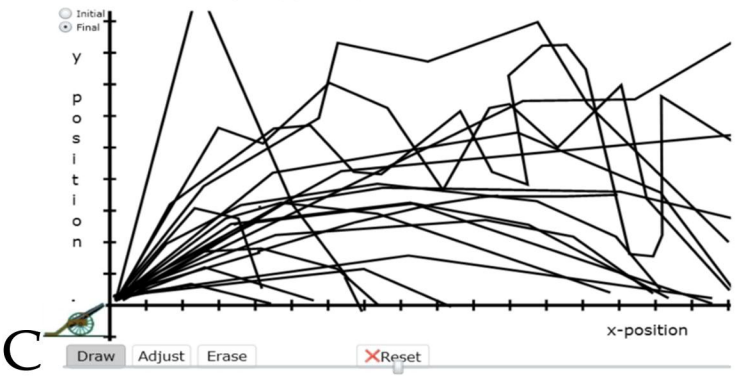
Draw a graph that describes the position of the cannon ball has moved as a function of time.



3.2: This is, of course, a macroscopic system. The cannon ball is very large compared to the molecules in the air. To make this activity a little more relevant to molecular systems consider how the system changes if the cannon is sitting at the bottom of a deep lake, and the cannon ball is fired into water rather than air. Draw the behavior of the cannon ball in water compared to its behavior in air (copied from the previous panel).



3.3: As you probably are already aware, molecules are very much smaller than cannon balls. Consider how the behavior of the cannon ball changes if it were the size of a molecule (fired from a molecular level cannon). Draw a fired molecule's trajectory (in water) as a function of time.



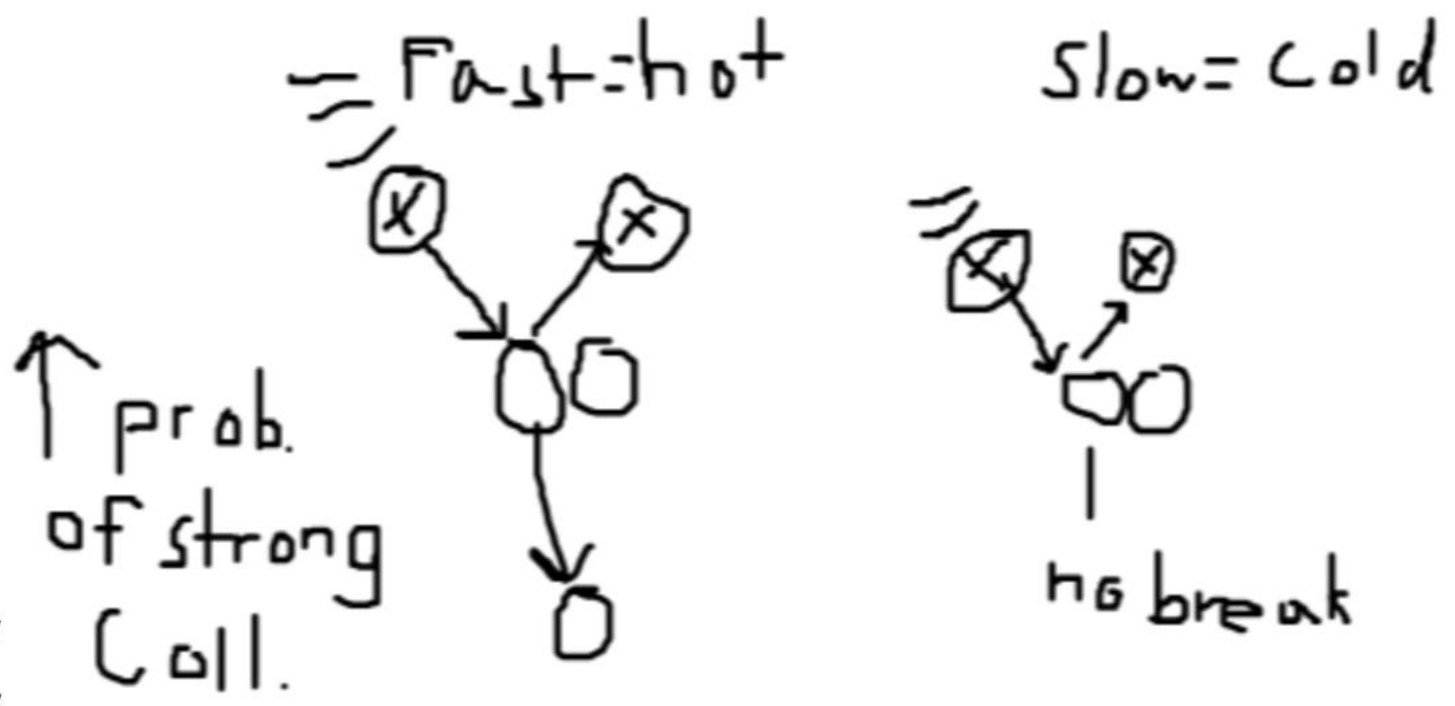
A



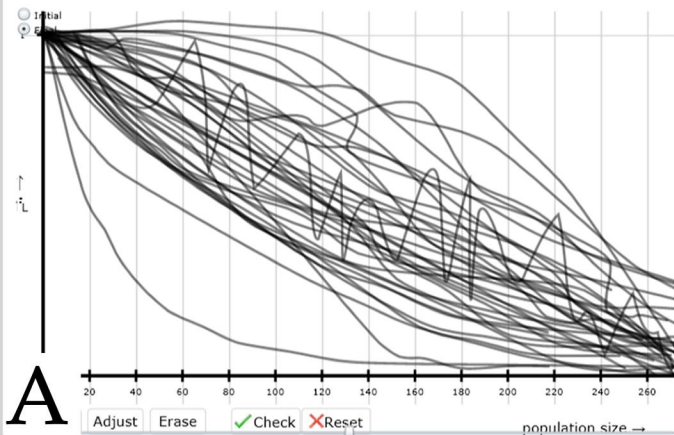
B



C

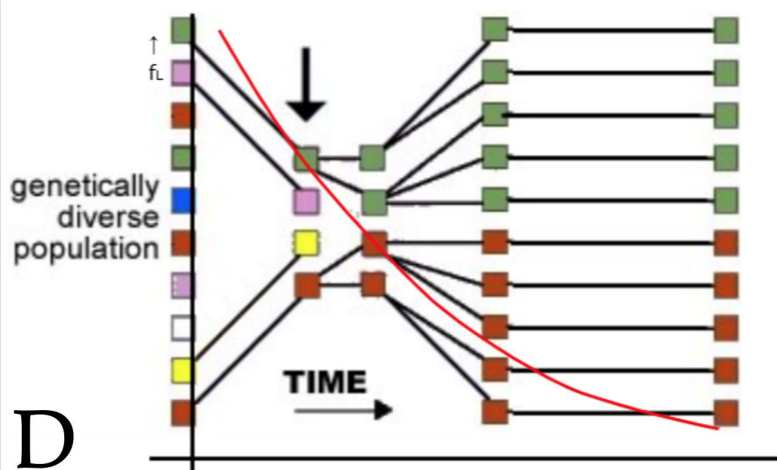
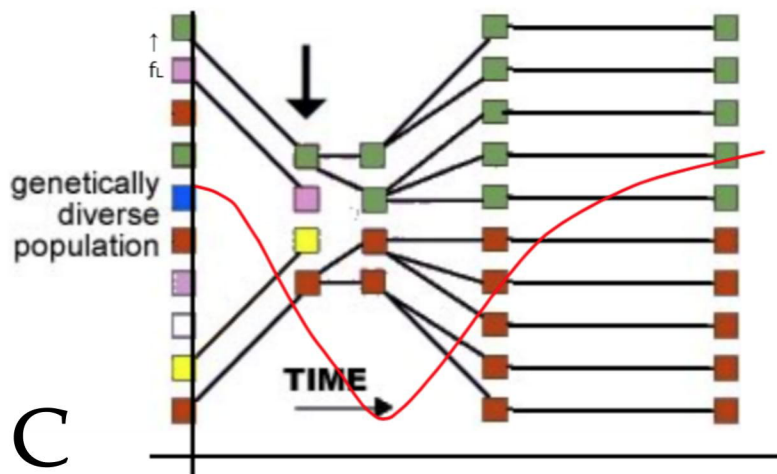
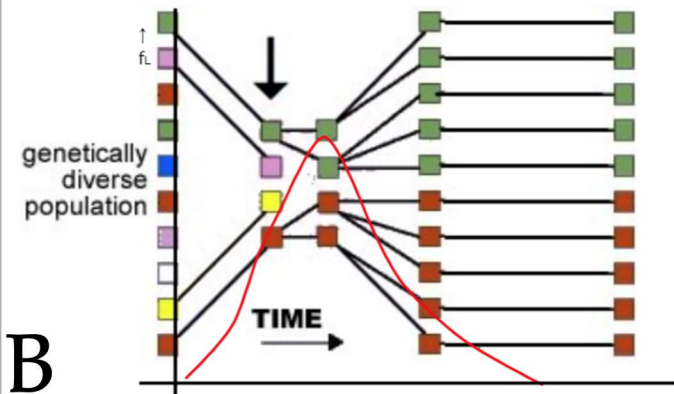


5.5: Draw a graph of the relationship between population size and f_L .



5.6: You begin with a genetically diverse population at $t=0$. A population bottleneck occurs (arrow) after which the population rapidly increases in size.

Draw a curve that represents the impact of genetic drift (f_L) as a function of time.



6.4: What is likely to happen if the concentration (number per cell) of TF molecules is reduced to less than 10. As before, assume that the inducer is added at $t=0$ and at a concentration high enough to activate all of the TF molecules. **Draw a graph** of the mature gene product as a function of time.

If you ran the same experiment 10 independent times, you would predict that ...

A

- it will be exactly the same each time
- it will be different each time
- it will be impossible to predict its behavior

B

- it will be exactly the same each time
- it will be different each time
- it will be impossible to predict its behavior

C

- it will be exactly the same each time
- it will be different each time
- it will be impossible to predict its behavior

time