Lecture 3:
CS 6306 / INFO 6306:
Advanced Human Computation

Today: Games with a Purpose (overt)

- Required readings:
  - Khatib, Firas, Seth Cooper, Michael D. Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, and David Baker. "Algorithm discovery by protein folding game players." *Proceedings of the National Academy of Sciences* 108, no. 47 (2011): 18949-18953.
  - Tuite, Kathleen, Noah Snavely, Dun-Yu Hsiao, Adam M. Smith, and Zoran Popović. "Reconstructing the world in 3D: bringing games with a purpose outdoors " In Proceedings of the Fifth International Conference on the Foundations of Digital Games, pp. 232-239. ACM, 2010.
- Additional readings:
  - Banerji, Manda, Ofer Lahav, Chris J. Lintott, Filipe B. Abdalla, Kevin Schawinski, Steven P. Bamford, Dan Andreescu et al. "Galaxy Zoo: reproducing galaxy morphologies via machine learning."*Monthly Notices of the Royal Astronomical Society*406, no. 1 (2010): 342-353.
  - Cooper, Seth, …. ”Predicting protein structures with a multiplayer online game.” *Nature* (2010).
  - *through a competitive game*.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
  - Lintott, Chris J., Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick et al. "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." *Monthly Notices of the Royal Astronomical Society* 389, no. 3 (2008): 1179-1189.
  - Kawrykow, Alexander, Gary Roumanis, Alfred Kam, Daniel Kwak, Clarence Leung, Chu Wu, Eleyine Zarour, Phylo players, Luis Sarmenta, Mathieu Blanchette, and Jérôme Waldispühl. "Phylo: a citizen science approach for improving multiple sequence alignment." *PloS one* 7, no. 3 (2012): e31362.
  - Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Gaikwad, S., Yoon, S., Treuille, A., Das, R., and EteRNA Participants, 2014. “RNA design rules from a massive open laboratory.” *Proceedings of the National Academy of Sciences*,*111*(6), pp.2122-2127.
  - Tuite, Kathleen, Nadine Tabing, Dun-Yu Hsiao, Noah Snavely and Zoran Popović. "*PhotoCity: training experts at large-scale image acquisition*

FoldIt

FoldIt

But first, “citizen science”

Citizen Science

- "scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions“ – *Oxford English Dictionary*
- History:
  - Often a pastime/hobby for the wealthy – “gentleman scientist”
  - Examples:
    - Astronomy: 1781 - William Herschel (while a musician) discovered Uranus in 1781
    - Archaeology: 1781 - Thomas Jefferson’s systematic excavation of a Native American burial mound in 1781
    - Meteorology: 1823 - Meteorological Society of London founded in 1823 “to receive meteorological observations from the cultivators of science…”
    - Meteorology: 1890 - US National Weather Service Cooperative Observer Program
    - Ornithology: 1900 - Audubon Society Christmas Bird Count
    - Entomology, botany, zoology, oceanography, etc.
    - Oxford English Dictionary: 1857
  - Development of “citizen sensor” approach

Citizen Science

- Today: Refers to both Internet-based efforts and local citizen science efforts (such as in schools)
- Three general modalities:
  - “Citizen sensor” – often on very large scales
  - STEM education
  - Collaborators/scientists

Prominent Internet-based Citizen Science Projects

- FoldIt:
  - Video game for protein folding
  - 300,000 players
  - Science outcomes:
    - George A. Khoury, … and Foldit Players. WeFold: a coopetition for protein structure prediction. *Proteins* (2014).
    - Christopher B. Eiben, … Foldit Players, …. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nature Biotechnology* (2012).
    - Firas Khatib, …. and Foldit Players. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences of the United States of America* (2011).
    - Firas Khatib, Frank DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, …. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* (2011).
    - Seth Cooper, … and Foldit Players. Analysis of social gameplay macros in the Foldit cookbook.*Proceedings of the 6th International Conference on the Foundations of Digital Games* (2011).
    - Seth Cooper, … and Foldit Players. Predicting protein structures with a multiplayer online game. *Nature* (2010).
    - Seth Cooper, … and Foldit players. The challenge of designing scientific discovery games. *Proceedings of the 5th International Conference on the Foundations of Digital Games* (2010).

Prominent Internet-based Citizen Science Projects

- FoldIt:
  - Video game for protein folding
  - 300,000 players
  - Science outcomes:
    - George A. Khoury, … and Foldit Players. WeFold: a coopetition for protein structure prediction. *Proteins* (2014).
    - Christopher B. Eiben, … Foldit Players, …. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nature Biotechnology* (2012).
    - Firas Khatib, …. and Foldit Players. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences of the United States of America* (2011).
    - Firas Khatib, Frank DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, …. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* (2011).
    - Seth Cooper, … and Foldit Players. Analysis of social gameplay macros in the Foldit cookbook.*Proceedings of the 6th International Conference on the Foundations of Digital Games* (2011).
    - Seth Cooper, … and Foldit Players. Predicting protein structures with a multiplayer online game. *Nature* (2010).
    - Seth Cooper, … and Foldit players. The challenge of designing scientific discovery games. *Proceedings of the 5th International Conference on the Foundations of Digital Games* (2010).

Prominent Internet-based Citizen Science Projects

- Galaxy Zoo:
  - Astronomical image labeling
  - 250,000 users
  - Science outcomes: 48 domain-science papers
  - Spawned the Zooniverse platform (46 domain-science papers)

Prominent Internet-based Citizen Science Projects

- EyeWire:
  - Mapping neurons in images
  - 200,000 players
  - Science outcomes:
    - M.J. Greene, J.S. Kim, H.S. Seung, the EyeWirers. Analogous convergence of sustained and transient inputs in parallel on and off pathways for retinal motion computation. *Cell Reports* **14**, 1-9 (2016).
    - J. S. Kim, M. J. Greene, A. Zlateski, K. Lee, M. Richardson, S. C. Turaga, M. Purcaro, M. Balkam, A. Robinson, B. F. Behabadi, M. Campos, W. Denk, H. S. Seung, and EyeWirers. Space-time wiring specificity supports direction selectivity in the retina. *Nature* **509**, 331-6 (2014).

Prominent Internet-based Citizen Science Projects

- Phylo:
  - Video game for multiple sequence alignment
  - Science outcomes:
    - Kawrykow, Alexander, Gary Roumanis, Alfred Kam, Daniel Kwak, Clarence Leung, Chu Wu, Eleyine Zarour, Phylo players, Luis Sarmenta, Mathieu Blanchette, and Jérôme Waldispühl. "Phylo: a citizen science approach for improving multiple sequence alignment." *PloS one* 7, no. 3 (2012): e31362.

Prominent Internet-based Citizen Science Projects

- Eterna:
  - Video game for RNA folding
  - “tens of thousands of players”
  - Science outcomes:
    - Anderson-Lee, Jeff, Eli Fisker, Vineet Kosaraju, Michelle Wu, Justin Kong, Jeehyung Lee, Minjae Lee, Mathew Zada, Adrien Treuille, Rhiju Das, and Eterna Players. "Principles for Predicting RNA Secondary Structure Design Difficulty." *Journal of molecular biology* 428, no. 5 (2016): 748-757.
    - Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Gaikwad, S., Yoon, S., Treuille, A., Das, R., and EteRNA Participants, 2014. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*,*111*(6), pp.2122-2127.

Science of Citizen Science: FoldIt, Eterna

- FoldIt:
  - Dun-Yu Hsiao, Seth Cooper, Christy Ballweber and Zoran Popović. User behavior transformation through dynamic input mappings. *Proceedings of the 9th International Conference on the Foundations of Digital Games* (2014).
  - Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Richard Snider, Jeff Lowdermilk, David Truong, Seth Cooper and Zoran Popović. The impact of tutorials on games of varying complexity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012).
  - Khatib, Firas, Seth Cooper, Michael D. Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, and David Baker. "Algorithm discovery by protein folding game players." *Proceedings of the National Academy of Sciences* 108, no. 47 (2011): 18949-18953.
- Eterna:
  - Treuille, A., & Das, R. (2014). Scientific rigor through videogames. *Trends in biochemical sciences*, *39*(11), 507-509.

Science of Citizen Science: Zooniverse

- Science Learning via Participation in Online Citizen Science, 2016.
- Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects, 2015.
- Designing for Dabblers and Deterring Drop-Outs in Citizen Science, 2014.
- Playing with science: gamised aspects of gamification found on the Online Citizen Science Project - Zooniverse, 2014.
- Why Won't Alien's Talk to Us: Content and Community Dynamics in Online Citizen Science, 2014.
- Volunteers' Engagement in Human Computation Astronomy Projects, 2014.
- Zooniverse: Observing the World's Largest Citizen Science Platform, 2014.
- "I want to be a Captain! I want to be a Captain!": Gamification in the Old Weather citizen science project, 2013.
- Creativity in citizen cyberscience: All for one and one for all, 2013.
- Learning by volunteer computing, thinking and gaming: What and how are volunteers learning by participating in Virtual Citizen Science?, 2013.
- Measuring the Conceptual Understandings of Citizen Scientists Participating in Zooniverse Projects: A First Approach, 2013.
- Galaxy Zoo: Motivations of Citizen Scientists, 2013.
- Galaxy Zoo: Morphological Classification and Citizen Science, 2012.
- Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers, 2010.

FoldIt

- Original Game:
  - Players twist proteins into shapes through direct manipulation tools
  - The quality of a shape is assessed by existing programs
  - Chemically important elements given visual cues, like colors
  - Technical terms replaced with more familiar words
  - Players are taught by advancing through levels

  - Why humans:

    “We hypothesized that human spatial reasoning could improve both the sampling of conformational space and the determination of when to pursue suboptimal conformations if the stochastic elements of the search were replaced with human decision making while retaining the deterministic Rosetta algorithms as user tools.”

FoldIt

- Assessment: New science outcomes:
  - Firas Khatib, Frank DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, …. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* (2011).

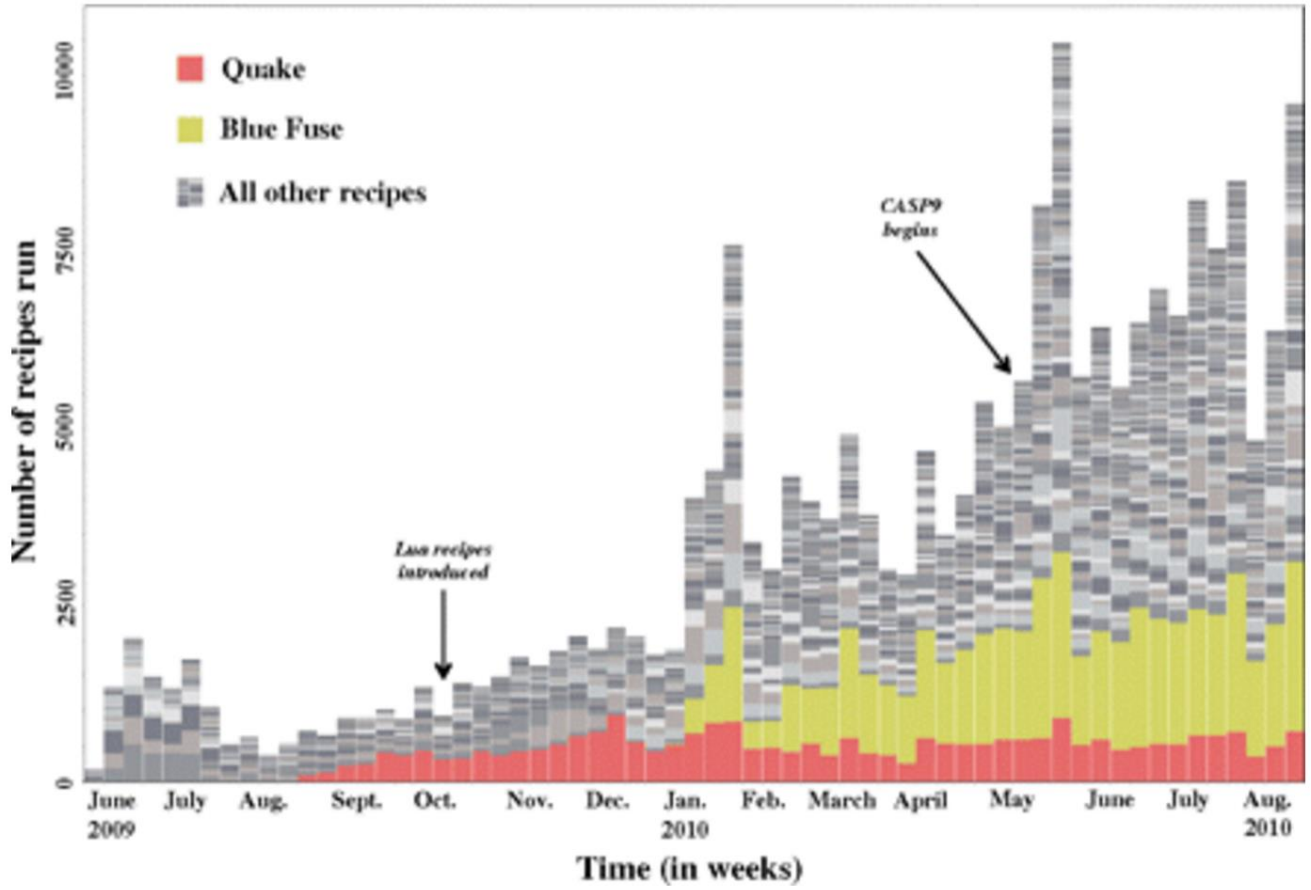
FoldIt

- Why should humans outperform computers?
  - “human ability to search over the space of possible strategies and adapt those strategies to the type of problem and stage of problem solving”
  - “The variability of tactics and strategies stems from the individuality of each player as well as multiple methods of sharing and evolution within the game (group play, game chat), and outside of the game”
- Learning how people solve these problems
  - Machine learning: Banerji, Manda, Ofer Lahav, Chris J. Lintott, Filipe B. Abdalla, Kevin Schawinski, Steven P. Bamford, Dan Andreescu et al. "Galaxy Zoo: reproducing galaxy morphologies via machine learning."*Monthly Notices of the Royal Astronomical Society*406, no. 1 (2010): 342-353.
  - Human computation: Khatib, Firas, Seth Cooper, Michael D. Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, and David Baker. "Algorithm discovery by protein folding game players." Proceedings of the National Academy of Sciences 108, no. 47 (2011): 18949-18953.

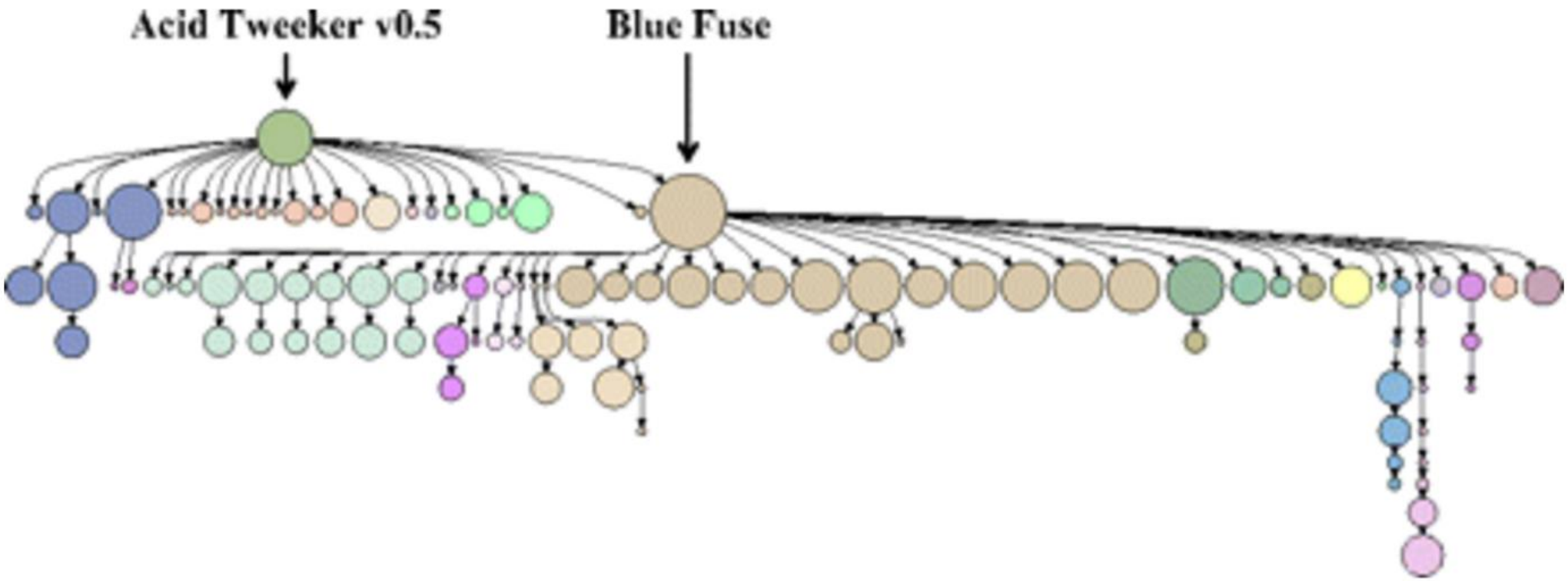
FoldIt

- Formulating algorithms – both science outcomes and assessment
  - Build off of FoldIt's ability to codify "recipes":
    - "We augmented standard Foldit play with the ability to create, edit, share, and rate gameplay macros, referred to as "recipes" within the Foldit game"
  - 568 players wrote 5,202 recipes
  - 721 Foldit players ran 5,488 unique recipes 158,682 times
  - Four categories:
    - perturb and minimize: add noise to avoid local minima
    - aggressive rebuilding: rebuild large regions (long run times)
    - local optimize: local minimization of a region
    - set constraints: adds constraints between residues or regions or tweaks secondary structure to influence optimization
  - Augment human effort by codifying low-level routines
    - Use always guided by human
    - Success requires human contributions

Assessment: History of Recipe Usage



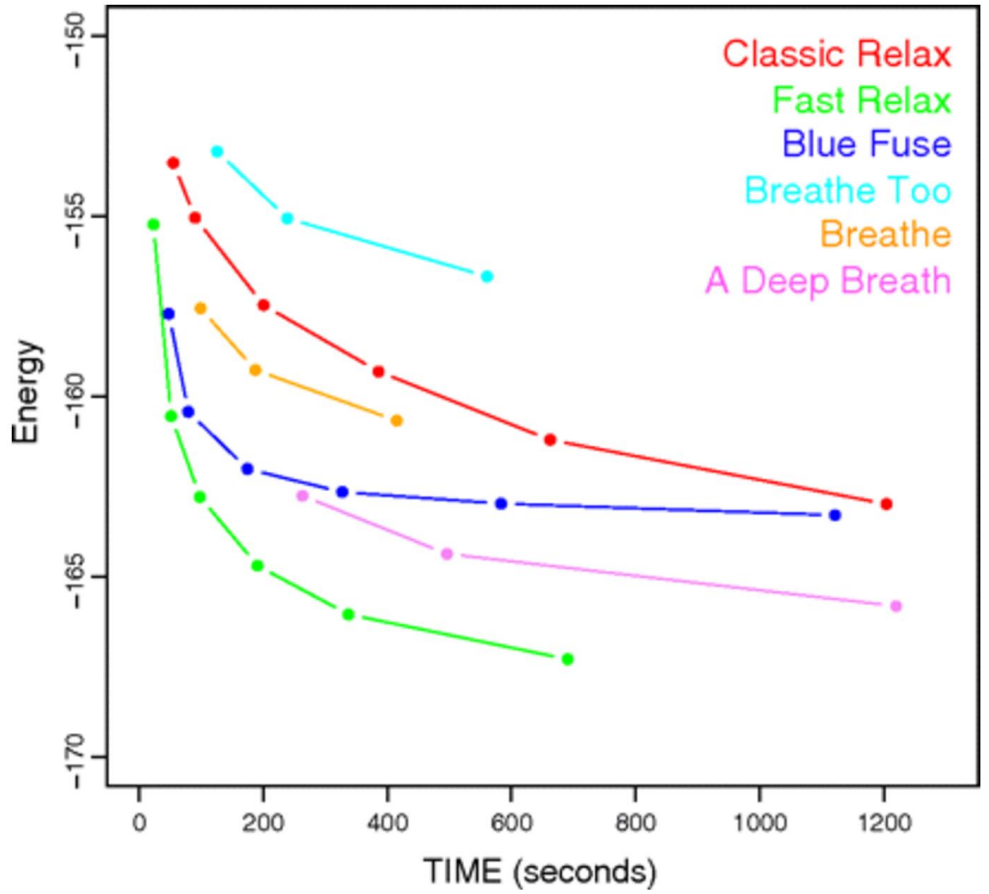
Assessment: History of Blue Fuse Recipe Creation



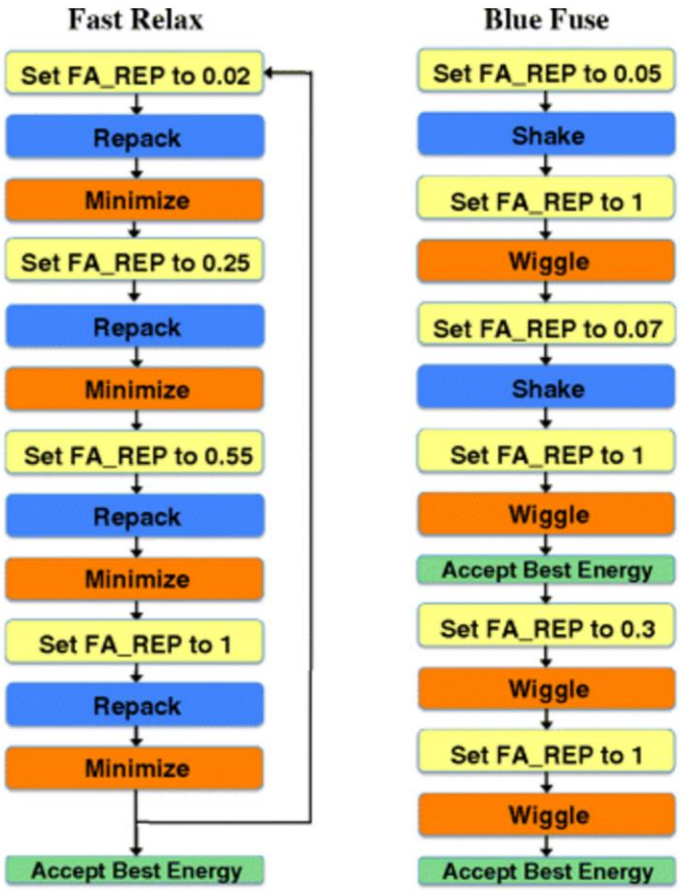
Assessment: Compared to Programs

- Same style developed independently by domain scientists
  - Levitt, M, Warshel, A (1975). Computer simulation of protein folding. *Nature* 253:694–698.
  - Levitt, M (1983). Protein folding by restrained energy minimization and molecular dynamics. *J Mol Biol* 170:723–764.

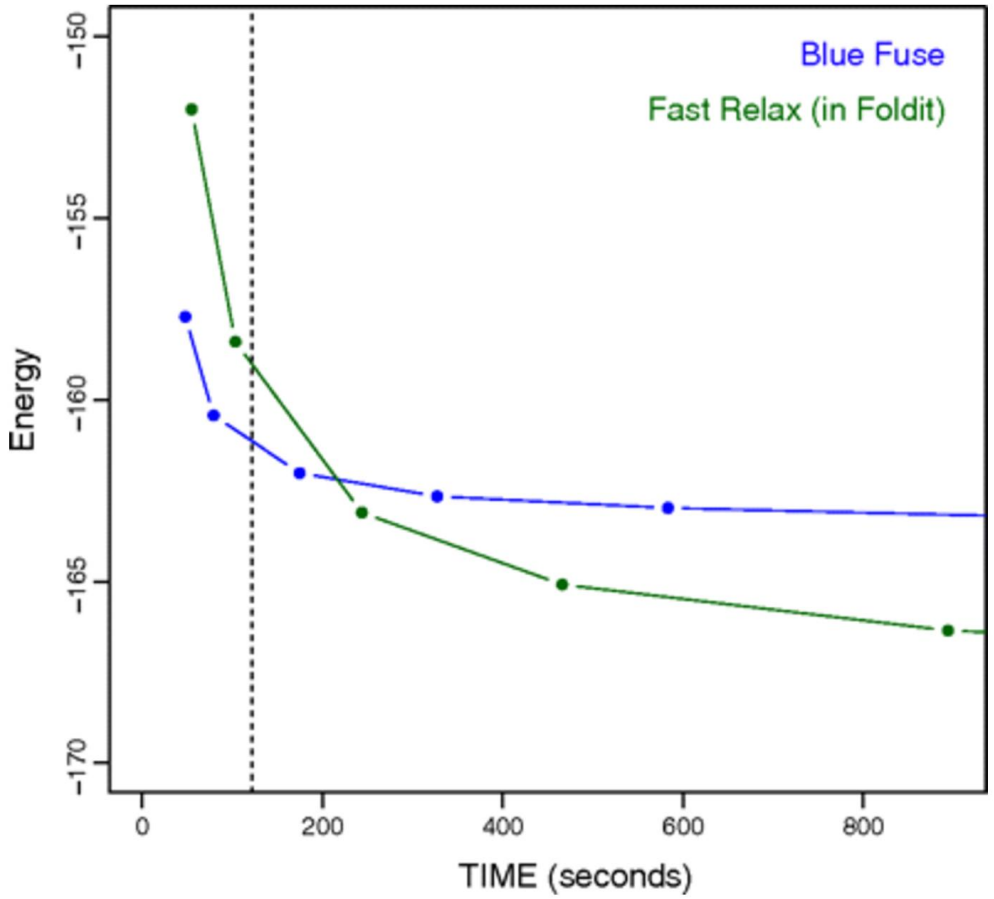
Assessment: Compared to Programs



Assessment: Compared to Programs



Assessment: Compared to Programs



Assessment

- Solved problems computers couldn’t
- Outcomes in use by players
- Comparable to programs and people

PhotoCity

- Motivation:
  - 3D reconstruction of scenes from photos requires photos (diverse, numerous) – 50 per building

PhotoCity

- Motivation:
  - 3D reconstruction of scenes from photos requires photos (diverse, numerous) – 50 per building

PhotoCity

- Goals:
  - Get players to take photos
  - Direct players to “gaps and fringes” *in a partial reconstruction*

- Only a few dozen players!

PhotoCity Game Play

PhotoCity Elements

- “Seed” reconstructions vetted by designers (20-200 photos)
- Models are dense point clouds (thousands of points)
- Photos add points to the model
- Players get points for taking pictures of flags:
  - Must overlap with existing points in the structure
    - Validates photos
  - Must overlap with an empty space
    - Adds to a model
  - Players get one point for each point added to the model
  - Capture flags at buildings by getting more points at the flag
  - Capture buildings by getting more points at the flag
  - (Upload photos via app or manually)

PhotoCity: Motivating Players

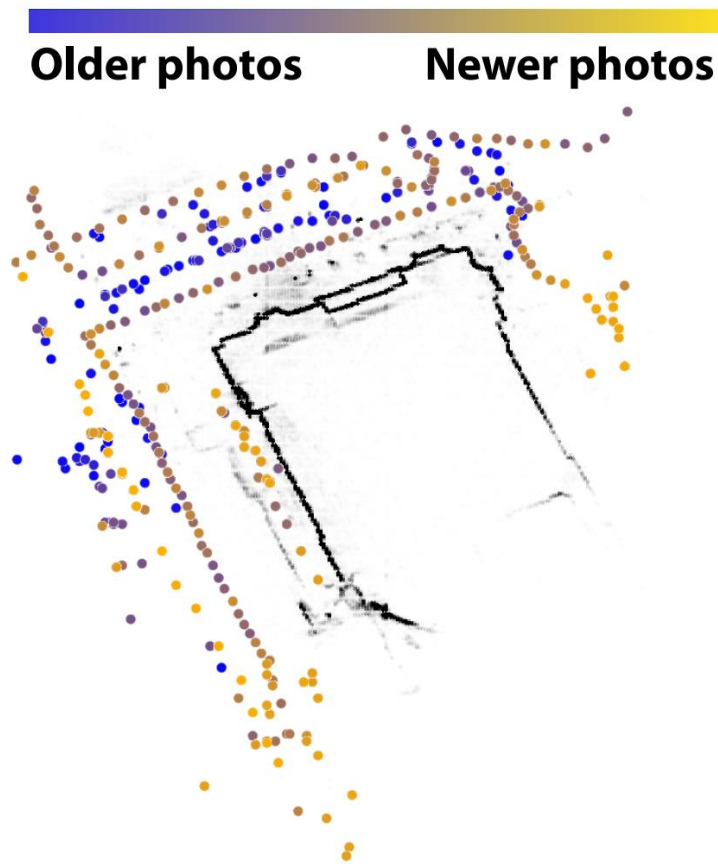
- Used theory from R. Bartle. *Designing Virtual Worlds*. New Riders Games, 2003 – 4 types of players:
  - Explorers
  - Achievers
  - Killers
  - Socializers
- Wanted to appeal to all of them
- Paid attention to rendering the game/models effectively

PhotoCity: Gaming Elements

- Teams:
  - Cornell vs UW
  - Red, green, blue, yellow
- Individual:
  - Leader boards
  - Titles (most models, flags, seeds, recruiting)
  - Prizes
  - Collectible gems (first player gets it)

PhotoCity: Outcomes

- 60% of photos are used in models
  (compared to 10% for Flickr-based project)
- 80% of photos from top 10 players
  - Top player responsible for 20% of photos
- Much denser coverage of campuses than Flickr
- Ratio of good/bad photos similar for “pro” and “non-pro” photographers
- Got to know campus


Older photos
Newer photos

PhotoCity: Outcomes

Model	Starting Photos	Starting Points	Ending Photos	Ending Points
CSE Front Entrance	58	18,131	515	148,189
Electrical Engineering	30	13,133	1,230	380,102
James J. Hill Statue	48	51,461	406	224,783
Southeast Corner of CSE	27	9,597	935	304,030
Fountain-facing Corner of EE	68	38,439	979	321,805
Commodore Apartments	41	9,973	581	66,777
Allen Library	92	28,945	840	131,558
Engineering Library	74	8,598	584	62,095
Guggenheim	101	85,239	1,394	622,355
Mechanical Engineering	46	27,022	741	273,889
Mary Gates (West Side)	32	36,891	1,488	761,777
Suzzallo Library	41	36,566	1,526	492,871
Hing Hay Park*	73	46,965	116	54,333

*Seed added by player during trial

Is this human computation?

Next Time: Programming Languages

- Required readings:
  - Barowy, D. W., Curtsinger, C., Berger, E. D., & McGregor, A. (2012). "Automan: A platform for integrating human-based and digital computation." *ACM SIGPLAN Notices*, *47*(10), 639-654.
  - Little, Greg, Lydia B. Chilton, Max Goldman, and Robert C. Miller. "Turkit: human computation algorithms on mechanical turk." In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pp. 57-66. ACM, 2010.
- Additional readings:
  - Kittur, Aniket, Boris Smus, Susheel Khamkar, and Robert E. Kraut. "Crowdforge: Crowdsourcing complex work." In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 43-52. ACM, 2011.
  - Patrick Minder, Abraham Bernstein, "CrowdLang: A Programming Language for the Systematic Exploration of Human Computation Systems." *Fourth International Conference on Social Informatics*, 2012.
  - Atsuyuki Morishima, Norihide Shinagawa, Tomomi Mitsuishi, Hideto Aoki, Shun Fukusumi. "CyLog/Crowd4U: A Declarative Platform for Complex Data-centric Crowdsourcing." PVLDB 5(12): 1918-1921 (2012)
  - Tosch, Emma, and Emery D. Berger. "Surveyman: Programming and automatically debugging surveys." *ACM SIGPLAN Notices*. Vol. 49. No. 10. ACM, 2014.
  - Tranquillini, S., Daniel, F., Kucherbaev, P., & Casati, F. (2015). "Modeling, enacting, and integrating custom crowdsourcing processes." *ACM Transactions on the Web*, *9*(2), 7.