Kyle J. Behymer, Infoscitex, USA
John M. Flach, Wright State University, USA

she ji
[ʃə: ʤi:]

From Autonomous Systems to Sociotechnical Systems: Designing Effective Collaborations

Abstract Effectiveness in sociotechnical systems often depends on coordination among multiple agents (including both humans and autonomous technologies). This means that autonomous technologies must be designed to function as collaborative systems, or team players. In many complex work domains, success is beyond the capabilities of humans unaided by technologies. However, at the same time, human capabilities are often critical to ultimate success, as all automated control systems will eventually face problems their designers did not anticipate. Unfortunately, there is often an either/or attitude with respect to humans and technology that tends to focus on optimizing the separate human and autonomous components, with the design of interfaces and team processes as an afterthought. The current paper discusses the limitations of this approach and proposes an alternative where the goal of design is a seamless integration of human and technological capabilities into a well-functioning sociotechnical system. Drawing lessons from both the academic (SRK Framework) and commercial (IBM's Watson, video games) worlds, suggestions for enriching the coupling between the human and automated systems by considering both technical and social aspects are discussed.

Keywords
Human-autonomy interaction
Collaborative systems
Human-machine teaming

Received June 6, 2016
Accepted September 8, 2016

Emails
Kyle J. Behymer
(corresponding author)
kbehymer@infoscitex.com

John M. Flach
john.flach@wright.edu

http://www.journals.elsevier.com/she-ji-the-journal-of-design-economics-and-innovation
http://dx.doi.org/10.1016/j.sheji.2016.09.001

Introduction

The goal of this paper is to argue that effective collaboration is critical to the success of human-machine teams, and to provide a framework (illustrated in Figure 1) for addressing the coupling between humans and machines such as autonomous agents. This is particularly important in the context of sociotechnical systems where multiple agents must collaborate to solve complex or wicked problems.[1] We will begin with a brief example to illustrate some of the dynamics of effective collaboration.

In 2005, Playchess.com hosted a chess tournament in which teams of human players could use computer assistance during matches.[2] The chess super computer Hydra was also entered into the competition, and after recently defeating Grand Master Michael Adams 5 ½–½ in a six game match, was considered to be the prohibitive favorite. Surprisingly, Hydra was eliminated before the semi-finals, with three of the four semi-finalists consisting of Grand Master-led teams equipped with supercomputers. Even more surprising was the fourth semi-finalist and eventual winner, team ZachS, composed of two relatively amateur chess players named Steven Crampton and Zackary Stephen using ordinary computers.[3]

The Elo Rating system[4] – a method of rating chess player skill level based on head to head results – puts team ZachS's victory into perspective. Table 1 lists Elo ratings ranging from novice to world champion. Current world champion Magnus Carlsen obtained the highest Elo rating (2882) in history for a human player (Garry Kasparov's best was 2851, Bobby Fischer's was 2785).[5] At the time of the tournament, Hydra's estimated Elo rating was 3000, and the runner up team was led by two 2600+ Grand Masters. Crampton and Stephen's Elo ratings were 1685 and 1398 respectively.[6]

Team ZachS was vastly outclassed in chess skill and computer hardware, yet overcame Hydra and the Grand Masters armed with super computers by quickly and efficiently manipulating their machines to deeply explore relevant positions and shrink the search space for their chess computers.[7] The higher skill level of Hydra and the Grand Masters equipped with super computers was not enough to overcome the seamless collaboration between the less skilled amateurs and their weaker computers. As Garry Kasparov stated, "Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process."[8]

Table 1. Chess Elo Ratings.

Elo	Skill Level
<1200	Novice
2000	Expert
2400	Master
2600	Grand Master
2700	World Champion

Synergy

In fact, the human-machine combination has the potential to outperform human-alone and computer-alone in many domains. For example, human forecasters at the National Weather Service can improve the accuracy of computer precipitation forecasts by 25% and computer temperature forecasts by 10% over computer-only forecasts,[9] and human-computer teams have the potential to outperform both doctors and computer algorithms at correctly interpreting mammograms.[10]

However, as the chess example illustrates, group performance is more than the sum of the abilities of the individuals that compose the group. For example, the collective intelligence of a group of people is more highly correlated with the group's social sensitivity, equality in turn taking, and the number of women in the group than the average intelligence of group members or the IQ of the group's

1 For discussion of 'wicked problems' see Richard Buchanan, "Wicked Problems in Design Thinking," *Design Issues* 8, no. 2 (1992): 14–19.

2 Various names have been proposed for this type of chess, including advanced chess, cyborg chess, centaur chess, and freestyle chess.

3 Clive Thompson, *Smarter Than You Think: How Technology is Changing Our Minds for the Better*, reprint ed. (New York: Penguin Books, 2014), 4–5.

4 The Elo Rating system, developed for chess by Arpad Elo in 1978, has also been used to measure skill level in many sports including baseball, basketball, football, soccer, and tennis.

5 Ratings from http://www.chessgames.com/chessstats.html.

6 "Dark Horse ZachS Wins Freestyle Chess Tournament," *Chess News*, last modified June 19, 2005, http://en.chessbase.com/post/dark-horse-zacks-wins-freestyle-che-tournament.

7 Garry Kasparov, "The Chess Master and the Computer," *The New York Review of Books*, February 11, 2010, accessed September 17, 2016, http://www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer/.

8 Ibid.

9 Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* (New York: Penguin Books, 2012), 125.

10 Mark Gaynor, George Wyner, and Amar Gupta, "Dr. Watson? Balancing Automation and Human Expertise in Healthcare Delivery," in *Leveraging Applications of Formal Methods, Verification and Validation. Specialized Techniques and Applications*, ed. Tiziana Margaria and Bernhard Steffen (Berlin: Springer-Verlag Berlin Heidelberg, 2014), 561–69.

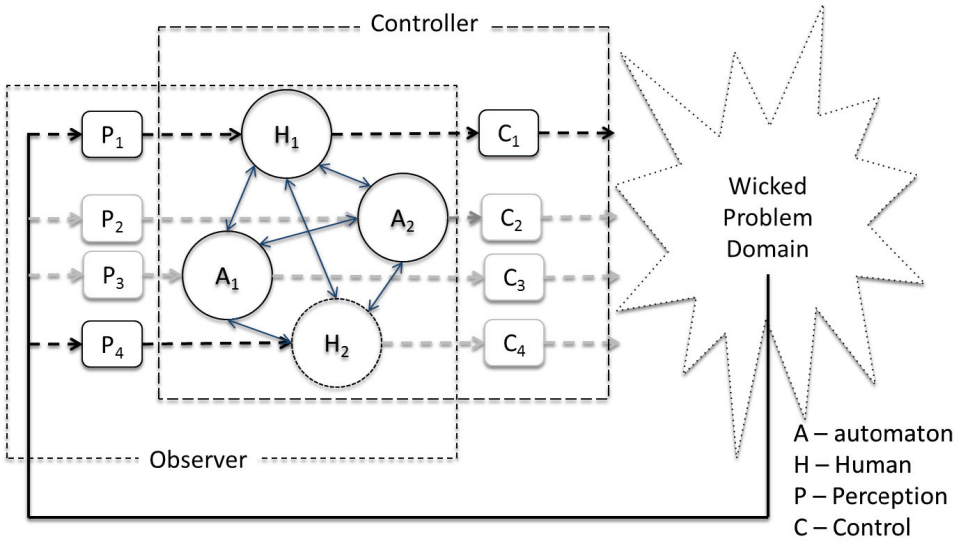


Figure 1 Framework for Human-Machine teams collaborating to solve a complex domain problem. Image © 2016 by Kyle J. Behymer and John M. Flach.

smartest person.[11] Similarly, cardiac surgery efficiency is more dependent on the surgical team's cumulative experience than the individual experience of the attending surgeon.[12] Teamwork beats talent when talent doesn't work as a team, as the saying goes. Pairing the best human with the best computer won't necessarily result in the best performance. Thus, there is much that the designers of autonomous technologies can learn from the literatures on team effectiveness[13] and distributed cognition.[14]

The human-machine team is like a pair of scissors cutting through the fabric of the work domain. Sharpening one or the other of the blades – increasing the capabilities of either the human or the machine – might lead to cleaner cuts. But if there were no screw – no effective interface – to hold the blades together, no matter how sharp they were, the scissors would not cut at all.

As shown in Figure 1, the distributed sociotechnical system includes multiple potential loops, where each loop is bounded in terms of access to information or perception (P) and action or control capability (C). The quality of the observer and control processes depends on the quality of coupling among the multiple loops. If the coupling is rich, then the sociotechnical system can be a better observer than any of the components. For example, each loop may provide unique information about the state of the problem domain – so redundant data across the loops can be useful in filtering sampling noise (averaging, common mode rejection). Also, rich coupling allows coordination of multiple actions to achieve a common goal. Without coupling, the actions of each loop will be a potential disturbance relative to other loops. If the coupling within the network of collaborating agents is rich – if there is effective communication – then the whole can serve as a more effective control system than any of the components. If the coupling within the network is poor, then there is a potential for the whole to be worse than the best component due to interference between loops.

Unfortunately, system developers often focus on increasing the capabilities of autonomous agents, without giving sufficient consideration to how they will interface with human operators. This approach often fails to recognize the technical limitations of autonomous components and the potential of a human-machine team. Additionally, focusing on improving the technical capabilities of autonomous agents without considering how these will interact with human operators often leads to poor coupling within the human-machine team. This paper proposes an alternative approach – frame the problem as interfacing to the problem domain *and* the other 'agents,' not only to improve observability and controllability, but also to

11 Anita Wooley et al., "Evidence for a Collective Intelligence Factor in the Performance of Human Groups," *Science* 330, no. 6004 (2010): 686–88, DOI: http://doi.org/10.1126/science.1193147.

12 Andrew Elbardissi et al., "Cumulative Team Experience Matters More than Individual Surgeon Experience in Cardiac Surgery," *Journal of Thoracic Cardiovascular Surgery* 145, no. 2 (2013): 328–33.

13 For example, see Eduardo Salas and Stephen M. Fiore, eds., *Team Cognition: Understanding the Factors that Drive Process and Performance* (Washington, D.C.: American Psychological Association, 2004).

14 For example, see James Hollan, Edwin Hutchins, and David Kirsh, "Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research," *ACM Transactions on Computer-Human Interaction* 7, no. 2, (2000): 174–96.

take the technical and social aspects involved in enriching the coupling between components into account.

The Prosthetic/Substitution Approach: The Technical Limits

Replacing the human user with autonomous systems, or at the very least, designing autonomous systems to either compensate for or overcome the limitations of a human user has been referred to as a prosthesis approach[15] or substitution-based approach,[16] and is based on the idea that designers should identify human weaknesses and replace them with automated strengths.[17] The origin of this approach can be traced to a 1955 Dartmouth manifesto in which a group of artificial intelligence scientists – John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon – proposed a goal of discovering how machines could solve the kinds of problems that had previously been the domain of skilled humans, without considering if/how these machines would interact with people.[18] The question wasn't how to design an autonomous system that could *collaborate* with a person to complete a task; rather, the question was how to design an autonomous system that could *substitute* for human capabilities.[19]

Advocates of the prosthetic/substitution approach often present humans as poor decision makers, citing studies in which human participants in contrived laboratory tasks perform poorly compared to mathematical decision-making models like Bayes' Theorem.[20] These studies conclude that human rationality is bounded and is therefore limited. What is often underrepresented is the fact that autonomous systems are bounded as well.

In 2005, the Defense Advanced Research Project Agency (DARPA) created a research program called COORDINATORS, whose goal was to develop hand-held automated agents that would help geographically distributed warfighters coordinate and adapt mission plans in response to unexpected events.[21] In 2008, a capstone exercise[22] was conducted to compare two different automated agent approaches – developed by two separate teams – with the performance of a control team of human operators.[23] One automated agent approach removed the human from the loop entirely; the team of humans it supported acted as actuators, only taking actions the agent assigned to them. This approach performed significantly worse than the control team of human operators, and the developers concluded that if automated agents are not provided with appropriate situation constraints, they will inevitably trend towards a subpar solution in the face of a highly dynamic environment.[24] The second automated agent approach fared slightly better, but only because its design allowed its developer team to input a human-devised strategy tailored to the specific scenario prior to the exercise. According to the designers of this approach, the automated agents still failed because they did not have an effective method of narrowing the enormous search space of the exercise's dynamic environment. The designers' key takeaway is worth quoting verbatim: "The most interesting result of the evaluation is that it is so difficult to outscore humans in a complex planning and scheduling problem."[25]

When an autonomous system is presented as the answer to the "problem" of human-bounded rationality it is inevitable that the technology will eventually reach its own limits. Consider Watson, the IBM supercomputer designed to compete in *Jeopardy!* – a television quiz show in which three contests compete to earn the most money by answering trivia questions. After two rounds, Watson was soundly defeating two of the best human *Jeopardy!* players ever, Ken Jennings and Brad Rutter,[26] by a score of $36,681 to $2,400 and $5,400 respectively. Then came *Final Jeopardy!* The category was U.S. Cities, and the clue was "Its largest airport is named for a World War II hero; its second largest, for a World War II battle."

15 Emilie Roth, Kevin Bennett, and David Woods, "Human Interaction with an 'Intelligent' Machine," *International Journal of Man-Machine Studies* 27, no. 5 (1987): 479.

16 Erik Hollnagel, "From Function Allocation to Function Congruence," in *Coping with Computers in the Cockpit*, ed. Sidney Dekker and Erik Hollnagel (Aldershot: Ashgate Publishing Company, 1999), 29–53.

17 Sidney Dekker and David Woods, "MABA-MABA or Abracadabra? Progress on Human-Automation Co-Ordination," *IEEE Intelligent Systems* 29, no. 5 (2002): 240–44.

18 Susan Epstein, "Wanted: Collaborative Intelligence," *Artificial Intelligence* 221 (2015): 36.

19 Gary Klein et al., "Ten Challenges for Making Automation a 'Team Player' in Joint Human-Agent Activity," *IEEE Intelligent Systems* 19, no. 6 (2004): 91–95.

20 John Flach and Fred Voorhorst, *What Matters?: Putting Common Sense to Work* (Dayton: Wright State University Libraries, 2016), 169–74.

21 Bob Kohout, "The DARPA COORDINATORS Program: A Retrospective," in *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems*, ed. Waleed Smari and Geoffrey Fox (New Jersey: IEEE, 2011): 342, also available at http://ieeexplore.ieee.org/document/5928708/.

22 While employed at JXT Applications INC., the first author managed a DARPA Phase II STTR (Small Business Technology Transfer) that developed the user interface used by one of the automated agent developer teams during this capstone exercise.

23 Rajiv Maheswaran et al., "Multi-Agent Systems for the Real World," in *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 2 (Budapest: AAMS, 2009): 1281–82.

24 Laura Barbulescu et al., "Distributed Coordination of Mobile Agent Teams: The Advantage of Planning Ahead," in *Proceedings*

Jennings and Ritter correctly answered "What is Chicago?", while Watson answered "What is Toronto?????"[27]

Watson's response elicited an audible groan from an audience full of IBM programmers likely thinking, "Toronto isn't a U.S. city." Except, as Watson was all too aware, Toronto *is* a U.S. city – in Illinois, Indiana, Iowa, Kansas, Missouri, Ohio, and South Dakota. Adding to the confusion, Watson was programmed to de-emphasize category names, as they are often only weakly tied to the content of the clue, and can contain puns or other forms of wordplay. So while this question was relatively easy for a human player, it proved to be Watson's Achilles' heel.

In their haste to replace irrational humans with rational machines, advocates of the prosthetic/substitution approach have failed to recognize that autonomous systems also have limits – what's more, they are overlooking a better solution. Imagine a warfighter teaming with a DARPA COORDINATOR agent. Imagine Brad Rutter teaming with Watson to play *Jeopardy!*. The critical point is that the rationality of all agents – human and machine – are bounded with respect to the complexity of many work domains. Thus, it will often be necessary to combine the capabilities of multiple agents, each with unique bounds and capabilities, in order to meet the demands for effective performance reflected in Ashby's Law of Requisite Variety.[28]

The Prosthetic/Substitution Approach: Unintended Social Consequences

> "John Henry hammered on the right-hand side. Steam drill kept driving on the left. John Henry beat that steam drill down. But he hammered his poor heart to death."[29]
>
> – Kennedy, Kennedy, and Baker, *Knock at a Star*

Thirty-five years ago, Weiner and Curry noted that the general public had two opinions in regard to automation: skepticism about its capabilities and fear of its consequences – widespread unemployment at best, and Orwellian dystopia at worst.[30] The prosthetic/substitution approach did little to alter these opinions in subsequent years, with coverage in the media being divided between fear mongering and disdain.[31] The story of John Henry's epic battle and ultimately Pyrrhic victory over the steam engine exemplifies the fear that people have of being replaced (or even destroyed) by automation, a fear that is omnipresent in popular culture. The first cinema robot appeared in Fritz Lang's 1927 silent film *Metropolis*, a *Maschinenmensch* (German for machine-human) created by the evil scientist Rotwang to replace Maria, an activist working to better the lives of the workers on whose backs the gleaming city of Metropolis has been built. The Maschinenmensch is designed to look exactly like Maria and has a single goal: to destroy Maria's reputation among the workers. The Maschinenmensch sows chaos among the workers and they riot, causing floods and destroying parts of the city. Eventually the subterfuge is discovered and the Maschinenmensch is burned at the stake. Similar themes are present in modern films such as *Terminator, The Matrix,* and *Avengers: Age of Ultron* – a machine designed to replace humanity turns on humanity. Both Schafer et al.,[32] and Parasuraman and Riley[33] have argued that these fictional portrayals have influenced society's perception of autonomous systems and may create dissonance when people interact with autonomous systems in the real world.

Another unintended social consequence of failing to take social factors into account when designing automated systems is disdain. In 1993, Microsoft started the Lumiere project, with the goal of developing an automated capability that could detect a user's goals based on their actions and provide assistance to the user to meet his or her goals.[34] In 1997, after more than 25,000 hours were spent on

of the 9th International Conference of Autonomous Agents and Multiagent Systems (AAMAS 2010), ed. Wiebe van der Hoek et al. (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2010): 1338.

25 Maheswaran et al., "Multi-Agent Systems," 1282.

26 Jennings won a record 74 straight games in 2004. Rutter has never been defeated by a human player in *Jeopardy!* and is not only the all-time *Jeopardy!* money winner but also the all-time game show money winner.

27 The number of question marks indicates the lack of confidence Watson had in its answer. See "What is Toronto?????" YouTube video, 1:38, from *Jeopardy!* on February 15, 2011, posted by Loginer, February 15, 2011, https://www.youtube.com/watch?v=7h4baBEi0iA.

28 W. Ross Ashby, *An Introduction to Cybernetics* (1956; Principia Cybernetica Web, 1999), 206, accessed September 17, 2016, http://pespmc1.vub.ac.be/books/introcyb.pdf.

29 X. J. Kennedy, Dorothy Kennedy, and Karen Lee Baker, *Knock at a Star: A Child's Introduction to Poetry* (New York: Little, Brown and Company, 1999), 21.

30 Earl Wiener and Renwick Curry, "Flight-Deck Automation: Promises and Problems," *Ergonomics* 23, no. 10 (1980): 996.

31 Epstein, "Collaborative Intelligence."

32 Kristin Schaefer et al., "The Future of Robotic Design: Trends from the History of Media Representations," *Ergonomics in Design* 23, no. 1 (2015): 13–19.

33 Raja Parasuraman and Victor Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no. 2, (1997): 230–53.

34 Eric Horvitz, "Lumiere Project: Bayesian Reasoning for Automated Assistance," accessed September 17, 2016, http://research.microsoft.com/en-us/um/people/horvitz/lum.htm.

usability testing, Clippy[35] was released as part of Office 97.[36] It was so despised that its removal from Office was included in the later Windows XP system sales pitch.[37]

Microsoft failed to realize how Clippy was perceived. "I HATED that clip. It hung around watching you with that nasty smirk. It wouldn't go away when you wanted it to. It interrupted rudely and broke your train of thought. It never actually had an answer to questions I had."[38] Microsoft spent 25,000 hours testing the technical capabilities of Clippy, but ignored the social components critical to ensuring a rich coupling between Clippy and the human user, dooming Clippy to failure.

Thus, in addition to considering the technical aspects related to the collaboration between humans and automated machines, it is also necessary to consider the social aspects. What does it mean for an automaton to be effective as a team player? How does an automaton earn the trust of an operator? How is it possible to strengthen the bonds among human and autonomous teammates? How can an automaton assist, without interrupting human processes or undermining human capabilities?

A Collaborative Systems Approach: Complementing Capabilities

> "Basically, meaningful interaction with an environment depends upon the existence of a set of invariate constraints in the relationships among events in the environment and between human actions and their effects."[39]
>
> – Jens Rasmussen, "Skills, Rules, and Knowledge"

If there is a rich coupling between the components outlined in Figure 1, the human-machine team will jointly bridge Hutchins, Hollan, and Norman's Gulf of Evaluation and Gulf of Execution[40] (see Figure 2A). However, if there is a poor coupling between the humans and technologies, then another gulf is introduced, creating additional uncertainties for each component (see Figure 2B).

On one side of these gulfs resides the human-machine team, with their goals and intentions. On the other side is the work domain. The size of the gulf of execution depends on the effectiveness of the actions or controllability the human-machine team has to achieve his or her goals. The size of the gulf of evaluation depends on how well the human-machine team can observe, perceive, and understand the state of the world in terms of their intentions – and find the best move on the chessboard. The goal-directed interaction between the human-machine team and the physical system is dependent upon constraints.[41] To span the gulf of evaluation it is vital to make the constraints of the work domain salient to the human-machine team.[42] To span the gulf of execution the constraints associated with the human-machine team – designing controls and displays that are consistent with a human operator's reasoning capabilities – must be considered.[43] Rasmussen's Skills, Rules, Knowledge (SRK) framework[44] defines three ways in which people represent constraints, signals, signs, and symbols which, in turn, distinguish three levels of human performance – skill-based, rule-based, and knowledge-based.

Skill-based behavior consists of sensory-motor tasks without conscious control.[45] In *rule-based* behavior, an individual has a set of predetermined solutions that are triggered by specific conditions, or signs. *Knowledge-based* behaviors occur when an individual encounters a novel unexpected situation for which no procedure exists.[46]

Consider a chef preparing a meal – chopping vegetables or continuously adjusting the gas flame of a burner to perfectly fry an egg are skill-based behaviors, and following a recipe is rule-based behavior. Now imagine that the recipe calls for vanilla extract – but when the chef looks in the pantry, the bottle is empty. The chef considers his or her options: (1) leave the cooking process, which is currently in a critical stage to acquire more vanilla extract; (2) skip the vanilla extract; or (3)

35 "Clippy" is a nickname; "Clippit" is its official name.

36 Eric Horvitz, "Lumiere Project."

37 Brian Whitworth, "Polite Computing," *Behaviour & Information Technology* 24, no. 5 (2005): 359.

38 Ibid.

39 Jens Rasmussen, "Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and other Distinctions in Human Performance Models," *IEEE Transactions of Systems, Man, and Cybernetics* 13, no. 3 (1983): 258.

40 Edwin Hutchins, James Hollan, and Donald Norman, "Direct Manipulation Interfaces," *Human-Computer Interaction* 1, no. 4 (1985): 319.

41 Kim Vicente, *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work* (Boca Raton: CRC Press, 1999), 10.

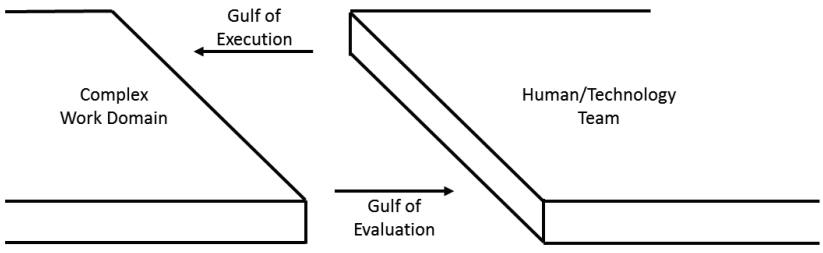
42 Kevin Bennett and John Flach, *Display and Interface Design: Subtle Science, Exact Art* (Boca Raton: CRC Press, 2011), 114.

43 Ibid.

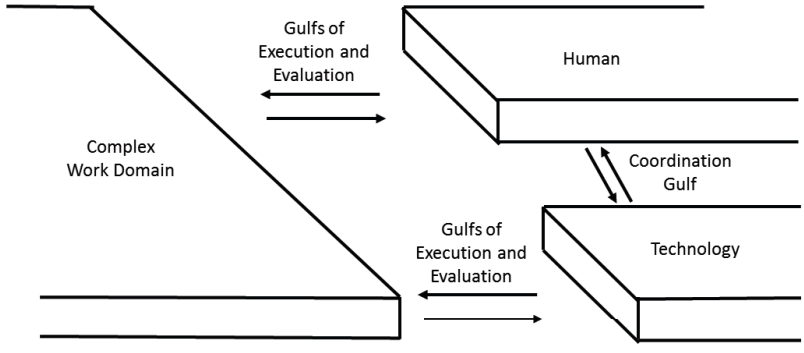
44 Rasmussen, "Skills, Rules, and Knowledge," 257–66.

45 Ibid., 258.

46 Ibid., 259.



A. Integration of Humans and Technology



B. Collection of Humans and Technologies

Figure 2 The gulfs of execution and evaluation. Adapted from Hutchins, Hollan, and Norman, "Direct Manipulation Interfaces" (see note 40).

find a suitable substitute. When the chef decides to try almond extract as a substitute, he or she is exhibiting knowledge-based behavior.[47]

The prosthetic/substitution approach attempts to replace the user at all three of these modes of interaction. While some researchers have suggested that autonomous systems are easiest to develop for skill-based behaviors and hardest for knowledge-based behaviors,[48] building autonomous systems that replace humans in all three categories can be very challenging. For example, cutting vegetables is simple for an experienced sous chef, but very difficult for an autonomous system.[49] Rather than replace the human, the focus should be on designing autonomous systems that support the human in all three modes of interaction – skill, rule, and knowledge-based.

Autonomous systems can be designed to augment a human's skills, rules, and knowledge (SRK) behaviors in several ways. First, the autonomous system should help the human explore the state space. IBM is currently repurposing Watson to help chefs do this very thing. Imagine a chef that has just returned from the garden with a plentiful bounty of sweet corn, lima beans, zucchini, and onions, and no idea of what to prepare. Using the Chef Watson app,[50] the chef can enter these ingredients and get back a variety of recipes featuring these ingredients – including zucchini tacos, zucchini fricassee, zucchini curry.

The chef can then collaborate with Chef Watson by adding constraints and narrowing the search space. For example, if the chef is in the mood for a specific type of cuisine, he or she can select the "Pick a Style" option. By default, Watson will recommend styles based on the ingredients – in this case Watson suggested Peruvian, Basque, Nuevo Latino, Moroccan, Tailgating, Tex Mex, Nashville, Cajun, and Israeli – but also provides the option for the chef to select an out of the box cuisine for these ingredients, like Japanese. Table 2 illustrates how Chef Watson has modified the zucchini taco recipe to infuse the dish with Japanese flavors. For example, Manchego cheese, which has a flavor profile similar to miso, has replaced goat cheese.

47 Note that if the almond extract replacement is deemed a success, the chef will likely switch to rule-based behavior in a similar situation in the future—"I am out of vanilla extract; I will use almond extract instead."

48 Mary Missy Cummings, "Man versus Machine or Man + Machine?," *IEEE Intelligent Systems* 29, no. 5 (2014): 66.

49 Ian Lenz, Ross Knepper, and Ashutosh Saxena, "DeepMPC: Learning Deep Latent Features for Model Predictive Control," in *Proceedings of Robotics: Science and Systems* (Rome: Sapienza University of Rome, 2015): 1–9, also available at http://www.roboticsproceedings.org/rss11/p12.html.

50 Available at https://www.ibmchefwatson.com/.

51 Joe Martin, "Top 10 Computer Game NPCs," bit-tech.net, last modified July 18, 2008, http://www.bit-tech.net/gaming/pc/2008/07/18/top-10-computer-game-npcs/5; Adam Dodd, "Top 10 Video Game Sidekicks," *Cheat Code Central*, accessed May 1, 2016, http://www.cheatcc.com/extra/top10videogamesidekicks2.html.

52 David Hodgson, *Half-Life 2: Raising the Bar* (Roseville: Prima Games, 2004), 153.

Table 2. Recipe differences (highlighted in gray).

Zucchini Taco	Japanese Zucchini Taco
Egg	Egg
Lima Bean	Lima Bean
Onion	Onion
Corn	Corn
Zucchini	Zucchini
Vegetable Oil	Vegetable Oil
Butter	Butter
Flour Tortilla	Flour Tortilla
Thai Chile	Chile de Arbol
Lemon Grass	Jalapeno Pepper
Pineapple Juice	Lemon Juice
Goat Cheese	Manchego Cheese
Coriander Seed	Caraway Seed

At this point, should the chef decide to make the Japanese zucchini tacos, Chef Watson has a recipe queued up and ready for the chef to access. And if the chef heads to the pantry and can't find any caraway seed, luckily Chef Watson has it covered – potential substitutes are generated upon request. The chef, aided by his or her sous chef, Chef Watson, can now get busy cooking.

The Chef Watson program illustrates how an automaton can be used both to complement cooking skills, and potentially stimulate a chef's creative ability to invent new solutions to the cooking problem. Note that the point here is neither to use Chef Watson as a substitute for a human, nor to use it to enforce procedural compliance by a human. Rather, Chef Watson becomes a creative partner – it helps the 'team' explore the cooking problem efficiently and effectively, think productively, and experiment with innovative alternatives.

A Collaborative Systems Approach: Creating Social Cohesion

Developers must also consider social aspects in order to facilitate a rich coupling between the human and autonomy. In 2004, Valve Corporation released *Half-Life 2* (*HL2*), the successor to their massively successful 1998 debut *Half-Life*. *HL2* is a first person shooter game in which gamers play the role of Gordon Freeman, a scientist who finds himself inspiring a resistance movement against a conquering alien force in a dystopian future. The *HL2* series deftly filled the shoes of its beloved predecessor by developing several innovative gameplay elements, including a fully realized AI sidekick named Alyx Vance. Alyx received almost universal praise, often ranking at the top or near the top of lists ranking the greatest non-playable characters of all-time, even years after her introduction.[51]

Alyx's success is a credit to Valve's intense development and play testing process. Over 100 actresses were auditioned to provide Alyx's voice, with developers seeking a voice actress that could be charming and warmly intimate, but could also be strong, confident, and believable.[52] During play testing it became clear that having Alyx be capable of providing assistance to the player wasn't enough. Developers initially tried to create a sense of urgency by having Alyx say things like "Hurry up!" and "Keep moving!" but players felt like they were being nagged,

and ended up hating Alyx. This led to a major design change – having Alyx almost always follow the player, rather than leading the way.[53]

Ultimately, making Alyx likeable was just as important – if not more so – than making her capable. The developers designed multiple scenes to humanize and endear Alyx to the player, and each scene had multiple variables that had to be just right. "If you don't like Alyx, you're not going to have much fun with Episode 1. So Alyx being likeable was one of our most crucial design goals. Little moments like the Zombine[54] joke are designed to make Alyx more endearing.… Surprisingly, lighting was really important too. Under red light, Alyx's self-deprecating groan looked more like she was sneering at the player for not getting the joke. Changing the lighting to blue and then adjusting the direction of the light so that it changed the shadows on her face fixed the problem."[55]

By the time the player reaches the end of *HL2: Episode 2* they will have spent many hours working with Alyx towards a common goal, and most players will have developed an emotional attachment to her, which Valve uses to devastating effect. At the end of *HL2: Episode 2* the player watches helplessly as Alyx's father is brutally murdered right in front of her. As Alyx – who is also restrained – screams in rage and agony, there's an incredibly brief moment when she glances back at the player, whispering "Gordon," her eyes pleading with the player for help. This gut-wrenching sequence continues as Alyx clings to her father's lifeless body, and her desperate sobs remain even after the screen fades to black.[56] This is how developers can create an AI that connects with the user on an emotional level.

Everything Valve got right with Alyx Vance, Microsoft got wrong with Clippy. Valve realized during the design process that Alyx should follow the player's lead and not control the action. Clippy would show up uninvited, take control of the user's mouse cursor, and keep coming back no matter how many times the user sent it away.[57] While Valve discovered that in the wrong lighting, what was intended to be a humanizing groan could be perceived as looking down on the player, Clippy's tone always seemed to convey that it knew better than the user. Valve spent thousands of hours perfecting Alyx's interaction with the player, resulting in one of the most beloved video game characters ever. Microsoft did not – and ended up with perhaps the most notorious automated assistant ever.

Conclusions

As many researchers have noted repeatedly,[58] effectiveness in sociotechnical systems will often depend on whether the technologies function as collaborative system team players. In many complex work domains, success is beyond the capabilities of un-aided humans, yet human capabilities are often critical to ultimate success. An important motivation behind the Cognitive Systems Engineering approach was the realization that no matter how carefully designed, all automated control systems will eventually face situations that were not anticipated at the time of their design.[59] Thus, at some point the human operators of those systems will be called upon to complete the design. In other words, the human operators will need to intervene to creatively deal with the requisite variety that was not anticipated by the designers of the automated systems. The SRK framework is specifically geared towards drawing attention to user interfaces, and ways to design representations so that human and automated systems can work together to creatively respond to the inevitable, unanticipated variability endemic to complex work domains.

In sum, the challenge is to move beyond an either/or attitude with respect to humans and technology – the classic "Humans are Better at/Machines are Better at" lists – that tends to focus on optimization of separate human and autonomous components as the top priority, and leaves the design of interfaces and team processes

53 "Half-Life 2: Episode One—Developer Commentary—Undue Alarm," YouTube video, 10:25, from Half-Life 2 Episode 1: Commentary by Matt Wood, posted by Grey Torch, April 13, 2015, https://www.youtube.com/watch?v=guAv4XhDGfc.

54 "Zombine" is a portmanteau of zombie and combine. The main antagonist force in Half-Life 2 is known as The Combine.

55 "Half-Life 2 Developer Commentary. Episode One [full, 2016 remake]" YouTube video, 50:45, from Half-Life 2 Developer Commentary. Episode One [full, 2016 remake]: commentary by Erik Wolpaw, posted by Steady Eddie, September 26, 2016, https://www.youtube.com/watch?v=118jCv8WT8g.

56 Scene available at https://www.youtube.com/watch?v=_zbrkhcF4_8 starting at 6:37.

57 Whitworth, "Polite Computing," 360.

58 Klein et al., "Ten Challenges," 91–95; Dekker and Woods, "MABA-MABA," 240–44.

59 John Flach, "Supporting Productive Thinking: The Semiotic Context for Cognitive Systems Engineering (CSE)," *Applied Ergonomics* (2015): 2.

as an afterthought. The alternative is to take a holistic perspective, and to begin thinking in terms of both/and, where the goal of design is a seamless integration of human and technological capabilities into a well-functioning sociotechnical system. Success in complex domains will ultimately depend on the ability of humans AND technologies working together as well coordinated teammates – each contributing unique abilities to create a team with the potential to be greater than the sum of its parts, and thus jointly bridge the gulfs of execution and evaluation in order to address the requisite variety of complex domains, or wicked problems.

Acknowledgments

The authors would like to thank Brian McKenna for comments on an earlier draft of this paper, as well as two anonymous reviewers.

Commentary

Value-Pluralism and the Collaboration Imperative in Sociotechnical Systems

Derek B. Miller, The Policy Lab®, and The Pell Center for International Relations and Public Policy, Salve Regina University, USA

derekbmiller@gmail.com

doi:10.1016/j.sheji.2016.12.001

In their article "From Autonomous Systems to Sociotechnical Systems: Design Effective Collaborations," Behymer and Flach make the case for the "seamless integration of human and technological capabilities into a well-functioning sociotechnical system."[1] The appeal is driven by the potential for increased effectiveness in a range of desirable actions that might result from better human-technology cooperation. The authors are also driven by a concern for the state of ongoing neglect. As they explain, there is currently too much attention paid to the "optimization of separate human and autonomous components" to systems thereby leaving "the design of interfaces and team processes as an afterthought."[2]

Should this appeal be heeded, it could result in a new commitment to an agenda of attending to this human-technology nexus, thereby enriching the possibilities for what they are calling sociotechnical systems. Importantly, though, the authors are not merely pointing to a contemporary gap in the human-technology working relationship, but instead are suggesting that the separation of technology from human support – full autonomy, as it were – is something of a chimera, because while technology can indeed be automated, it cannot by its nature respond to the range of experience that would make it function well in all circumstances:

> "No matter how carefully designed," they argue, "all automated control systems will eventually face situations that were not anticipated at the time of their design. Thus, at some point the human operators of those systems will be called upon to complete the design. In other words, the human operators will need to intervene to creatively deal with the requisite variety that was not anticipated by the designers of the automated systems."[3]

On that basis, it is essential that the human-technology team process and interface not be treated as an afterthought, but addressed specifically.

What I would like to do, here – albeit briefly – is provide support to Behymer and Flach's appeal for an enriched agenda on this matter by situating the problematic they observe in a wider context, and demonstrating that not only will automated control systems eventually face these complex situations, but humans will too. And furthermore, we always will.

Today, at the time of writing, the international community is engaged in a sophisticated and urgent discussion around the topic of lethal autonomous weapons. A forum for the advancement of this discussion is the Convention on Certain Conventional Weapons (CCW), and the Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), information about which can easily be found online through the United Nations Office for Disarmament Affairs, and