The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
AI, Ethics, and Society: Technical Report WS-16-02

Playing Games Across the Superintelligence Divide

Aaron Isaksen,* **Julian Togelius,*** **Frank Lantz,†** **Andy Nealen*†**

*NYU Game Innovation Lab †NYU Game Center
New York University
2 MetroTech Center
Brooklyn, NY 11201

Abstract

Humans may one day create *superintelligence*, artificially intelligent machines that surpass mankind's intellect. Would these artificial intelligences choose to play games with us, and if so, which games? We believe this question is relevant for the ethics of general AI, the current widespread integration of AI systems into daily life, and for game AI research. We present a catalog of scenarios, some good for humanity and some bad, in which various kinds of play might take place between humans and intelligent machines. We assume a superintelligence, because of its greater cognitive ability, would stand in a similar relation to us as an adult does to a child, an expert to a novice, or a human to an animal. We define friendly games, learning games, observational games, and domination games, and proceed to consider games adults play with children, experts play with novices, and humans play with animals. Reasoning by analogy, we imagine corresponding games that superintelligences might choose to play with us, finding that domination games would pose a significant risk to humanity.

Introduction

This paper poses the question of what kind of games that superintelligent artificial intelligences might choose to play with us. By *superintelligent artificial intelligence* or *superintelligence*, we mean a computer system that significantly surpasses general human cognitive ability in all or most respects (Good 1965), with the ability to "learn, reason, and plan...across a wide range of natural and abstract domains" (Bostrom 2014). In particular, such a system would be better than us at playing practically all games in the cognitive domain and solving general gamelike cognitive tasks. Today's computers can already exceed our abilities in specific games like Chess (Newborn 1997) or Checkers (Schaeffer et al. 2007), but these programs only do well at one game at a time and are thus not generally intelligent. In contrast, a superintelligent AI would exceed our abilities in effectively all games that humans play.

Games are an essential part of human society, with play deeply integrated into human existence (Huizinga 1949; Caillois and Barash 1961; Sutton-Smith 2009). The exploration and better understanding of how, what, and why we play is a question as old as mankind, and there is no expectation that interest in games will decrease in the near future. With the proliferation of computers, real-time graphics, mobile devices, and the ubiquity of the internet, games have become so popular that it's likely "the 21st century will be defined by games" (Zimmerman and Chaplin 2013).

Games have also long been associated with research in artificial intelligence, at least since Turing formulated his eponymous test as a game (Turing 1950). Game-based competitions have been used as AI benchmarks for as long as the field has existed, starting with classic board games and in the last decade expanding into video games. In recent years, there has been increasing recognition that benchmarks of general AI need to be based on more than a single game (Schaul, Togelius, and Schmidhuber 2011), as reflected in the General Game Playing Competition (Genesereth, Love, and Pell 2005) and the General Video Game Playing Competition (Perez et al. 2015), where AI players are tested on their ability to play unseen games using dynamic learning of strategies.

Given the importance of games to both society and artificial intelligence, we believe it's likely that teams building general intelligence systems will be interested in using games and game learning, whether that team operates in the academic, defense, financial, or other domain (Barrat 2013).

Our strategy for exploring this topic is primarily to reason from analogy. To get a sense of the types of games superintelligences might play with humans, we investigate which games humans play with other living things that have a lower order of intelligence or significantly different level of skill: namely *Adult-Child*, *Expert-Novice*, and *Human-Animal* play.

We believe this research has benefits for the present, even though we are most likely far from reaching superintelligence. Firstly, there are ethical concerns with artificial intelligence which we believe should be more well known, and this paper is an attempt to raise these concerns within the AI and game research community. Secondly, as game designers and AI researchers, we aim to make better games and better game-making systems, and an exploration of how humans might interact with players that are significantly more intelligent than us could lead to new areas for game design research and development. Finally, it suggests interesting research problems for artificial intelligence as applied to games, and offers us some understanding of how research in our field of game design and game AI might benefit the development

	Adult-Child	Expert-Novice	Human-Animal
Friendly	*Peek-a-boo, Tea Party, Hide & Seek, Building Blocks*	*Handicapping in Go or Golf*	*Hamster Wheels, Dog Toys, Polo, Racing*
Learning	*Candy Land, Tic-Tac-Toe, Tee-Ball, Chutes & Ladders, Trivial Pursuit for Kids*	*Tutoring, Mentoring, Coaching*	*Dog Tricks*
Observational	*Playground Park Bench, Spectating Kids Sports, Psychological Studies*	*Spectating Professional Sports*	*Birdwatching, Dog Park, Fish Tank*
Domination	*Forcing Play, Tickling*	*Player Killing, Griefing*	*Bull Fighting, Fishing, Hunting, Rodeo, Biotic Games*

Table 1: Our categorizations of games which are played by players of different levels of intelligence. Each column represents a different category of relative intelligence between the players. Each row represents the inter-player relations inherent in their core interactions. Each cell contains some of the example games discussed in detail in this paper.

of general artificial intelligence, and protect from its ethical dilemmas and existential dangers.

Even if one does not believe that superintelligence is possible, or is very unlikely, much of our discussion here also applies to systems and machines with high-functioning artificial intelligence that are beyond our understanding and explicit control. Many AI systems already today have a significant impact and controlling influence on humans (Bryson, Kime, and Zürich 2011). Although these machines may never have human-level self-conciousness or actually desire to do something in the ways that humans desire things, machines still have an impact on our lives and how we play games. For the purposes of this paper, we hope a superintelligence skeptic can mentally replace the terms "superintelligence" with "highly-functioning and deeply-integrated artificial intelligence" and find our arguments still valid.

In this paper, we do not take a stance on how to build superintelligence, when it could be done, or even whether it is in principle even possible. Current theories on superintelligence rely on the idea of a *seed AI*, created by humans, which has the ability to learn for itself, self-educating and self-modifying until it far exceeds human abilities and becomes superintelligent (Goertzel and Pennachin 2007). Here we simply assume that superintelligence could be achieved and try to reason about what games such minds might play with us. It is debatable if there will be a singleton superintelligence or multiple superintelligences (Bostrom 2006), but for this paper we will assume there will be more than one, as it maps better to our own experience with games. We do not assume that superintelligent AI will always be friendly towards humans (Yudkowsky 2001), that the future is always positive (Fox and Shulman 2010), or make any assumptions about nanotechnology or the Singularity (Vinge 1993; Kurzweil 2005). We will show that some of these games could be a blessing for humans, while others pose severe dangers. This implies that people working on advanced AI systems, which might someday work on a seed AI, should consider the impact if a higher order intelligence were allowed to play any game it chose without ethical guidelines.

Categorizing Games and Play Between Different Levels of Intelligence

It's very hard to reason about how we might interact with agents whose cognitive skills are far superior to our own. By looking at examples of analogous situations – *Adult-Child* play, *Expert-Novice* play, and *Human-Animal* play – in which *we* are the smart ones, and then imagining a more complex version of that situation and picturing ourselves on the other side of the table, we may gain intuitions and insights that are grounded in something beyond pure speculation.

These example games are categorized according to the inter-player relations inherent in their core interactions: friendly, learning, observational, and domination. *Friendly games* are played for the mutual enjoyment of all parties, and usually with voluntary participation. *Learning games* are those where the main purpose is for the adult or expert to teach some particular skill or behavior. In *observational games* one player takes no active part (except perhaps as "prime mover" in setting up the limits of interaction) but is an engaged spectator of the play experience. *Domination games* are those where one party exerts dominance over the other, such that the dominated player is not enjoying the experience and may have no choice but to participate, and where the dominating player may be inflicting mental or physical pain and suffering (and possibly even death in the case of some games with animals). Table 1 shows examples of games discussed in this paper based on our cross-categorization. The boundaries between these categories are not precise, and some games can fit into more than one category.

For each type of game, we try to find analogues that would fit the relationship between humans and superintelligences, where adult humans would switch roles and become the cognitively inferior player. It quickly becomes apparent that some of these games are more desirable than others from our human perspective. Some of the games and modes of play we describe could to some extent be implemented with today's domain-specific artificial intelligence techniques, whereas others would require true superintelligence.

Would Superintelligences Play?

Before we explore in detail the list of example games, it is worth considering whether superintelligences would choose to play games at all, and if they play games, whether they would play games with *us*. At least one model of a computer having fun while playing games has been presented previously (Togelius, Shaker, and Yannakakis 2014). It could be argued that minds that are so much more intelligent than us would have advanced beyond the need for play and games. Or perhaps, if they would indeed play, they would choose to play with each other rather than with cognitively inferior humans. This stands in contrast to today's Chess computers which do not have a choice whether or not to play.

Nonetheless, we believe there is a significant probability that superintelligences would still play games. While this topic is worthy of more attention, we will give a brief sketch of several reasons why superintelligences might play.

Playing by Design. It is possible to imagine forms of AI that are superintelligent and nonetheless bound by behavioral constraints that were implemented into the seed mind. In some cases we may program superintelligent AI with utility functions that explicitly cause them to play games, or that drive them to desire to play games. The superintelligence may also have evolved from a successful general game playing system, such that it maintains its desire to play games from its ancestor versions.

Inheriting the Play Instinct. In those scenarios in which superintelligent AI start out as emulations of human brains and then achieve superintelligence through self-modification (Sandberg and Bostrom 2008) the advanced AI may retain many of the underlying motivations, preferences and behavioral tendencies of the human "brain seed". Superintelligences of this type may posess a desire to play games that resembles an extruded or extended version of our own, in the same way that we possess behavioral traits that resemble those of the animals from which we evolved.

Developmental Play. A large variety of animals play, including most mammals and many species of birds (Alcock 1993). Examples range from cats playing with captured mice to Alaskan ravens "snowboarding" on steep snow-covered roofs to dogs play-fighting in the park. Engaging in playful activity might be associated with higher survival fitness, meaning that animals with a capacity and propensity for playfulness have a higher fitness for propagating their genes. A key benefit of play is that it can allow animals to train useful skills in relative safety. For example, play-fighting mimics real fighting without danger, and raven-snowboarding might provide training in various motor skills. The proximal motivation for engaging in play is almost certainly not the learning (the animal likely does not even know it's learning) but the enjoyment gained from playing; evolution seems to have favored animals that enjoy learning this way. There are also other explanations for why animals play. For example, David Graeber argues that all animals engage in play and that play is ontologically primal; this, however, is not a mainstream view (Graeber 2014).

Explanations for why *humans* play are much more diverse (Huizinga 1949; Caillois and Barash 1961; Sutton-Smith 2009), and this section does not attempt to give a fair overview of them. It is worth noting that many of them include the learning effects of play and games as a key motivator. Vygotsky postulated "proximal zones of development" consisting of tasks which a child could not perform unaided, but which the child could perform partially or with aid. The child seeks out these tasks for self-directed play, as they allow for most efficient learning (Vygotsky 1967). Current game design thinking also emphasizes the role of learning in playing, and in particular how some of the pleasurable aspects of play may derive from evolutionarily acquired rewards that encourage learning. According to Koster, a game is fun only as long as the player learns from playing it; a trivial or impossible game is not worth playing (Koster 2005).

Would these motivations for play carry over to an advanced AI? Would an artificial system that seeks to improve its own cognitive abilities play to learn in the same way a human or animal does? One relevant model here is Schmidhuber's formal theory of fun, curiosity and creativity. According to this theory, any system (natural or artificial) that seeks to maximize its learning rate will try to seek out those environments where it can learn as quickly as possible, which in general are those areas which it has only partial mastery (Schmidhuber 2010). These essentially correspond to Vygotsky's proximal zones of development. Schmidhuber's general (but AI-centric) model and Vygotsky's human-centric theory both point out the positive effects generated by exploring such environments and tasks. Moreover, exploring artificial problems would allow learning without associated risks. From some perspectives, a risk-free environment with tasks of a sufficient difficulty that allow a high amount of learning can simply be called a game. Thus we have at least one plausible reason why a self-improving AI would choose to play games.

Playing with Us. Finally, why would an artificial intelligence choose to play with humans? In the case of designed play, we may simply compel them to by embedding the motivation to play deep within their code. But there are reasons that intelligences who are not constrained in this way may still wish to play with us. To start with, humans will be the most advanced general intelligences on the planet, save the superintelligences themselves. Whether or not the superintelligences are strictly patterned on human minds, we are likely to be sufficiently different from them to be worth exploring and studying. It is worth stressing that though the superintelligences are more intelligent than we are, they may still be as stumped in trying to predict our behavior in detail as we frequently are when observing much simpler organisms. Another reason they might attempt to understand us in more detail is that we gave rise to them, either through creating a seed AI, a template for whole-brain emulation, or by designing them in their entirety.

There might still be other reasons for why they would like to play with us—maybe they would love us or care about our welfare? But no specific assumptions are required for the following sections beyond the general notion that play between humans and superintelligent AI is possible.

Games Between Adult and Child

Adults play games with children to teach and for the enjoyment of all involved. Even though the adult and child are mismatched in ability, as it's likely that an adult would win at strategic games against a child, adults can still play at a level for the child to have a good experience. We refer to children as having a lower order of intelligence than adults not in terms of their potential, but in terms of their relative level of cognitive skill and general game playing ability.

Our analogy here maps the adult to superintelligent AI and the child to humans; this puts humans in a relationship where the machines are providing for our needs and are interested in our safety, happiness, and cognitive development.

Adult-Child Friendly Games

In our categorization, friendly games between adult and child are games that adults setup for the pure enjoyment of the child and themselves. The point of these games isn't necessarily to teach children specific lessons, although all games do have some aspect of cognitive development since players are always learning while playing (Prensky 2005).

Adult-Directed Play. This category is for games played with children where the adult is driving the play experience while the child is reacting. Games like *Peek-a-boo* or *Scary Monster* (where a parent acts out the role of a mock threat) fit into this category, and simply provide joy and laughter to the child and adult. More complex games of this type, organized and run by adults, might be played in groups, such as *Easter Egg Hunt*, *Pin the Tail on the Donkey*, or *Musical Chairs*.

Superintelligence-Directed Play might show up in a form where the AI is providing playful activites to entertain us, as is currently done by specialized AIs in procedural content generation systems (Togelius et al. 2011). This might be in the form of games that the AI suggests for us to play with each other, or could be a type of enhanced playful reality where games are integrated into our everyday lives. For example, in a future where all vehicles are controlled or at least overseen by AI, a superintelligence might choose to invent races or high-speed car chases to entertain people who still like to drive (these would be completely safe, as the AI would take over if anything dangerous was about to happen). There could be vast, world-spanning AI-created treasure hunts, complex, multi-layered obstacle courses, or even giant, city-devouring monsters we must work together to defeat. Like the parent playing the role of Scaryt Monster we remember from our childhoods this would be an experience that combines terror and comfort to produce a peculiar form of glee.

Child-Directed Play. Child directed play is when the child creates the game and then invites the adult to play along. This may be something like *Playing Tea Party*, where the child makes up a fictional social occasion and assigns a specific role for the adult to play, or *Building Blocks*, where children assemble structures which are then shown to the adult for approval. Also in this category would be when a child modifies an existing board game with new rules and invites the adult to play this new version.

The Superintelligence analogy here might be scenarios in which the AI fulfills a specific role within a human-created game: performing a character, filling in the details of a sketched-out design, judging human performance, playtesting our games with us, etc. Specialized AIs are already an important part of constructing narratives, character behaviors, and content for games (Mateas and Sengers 2003; Young et al. 2004).

Adult-Child Co-Play. These games are played together with adults and child both playing necessary roles but neither fully driving the game. Some examples include playing on a *Swing Set*, *Kicking a Ball*, *Hide & Seek*, or *I Spy*.

A superintelligence playing together with us could clearly beat us in all competitive cognitive games, but it still might play with us at an appropriate level just to keep us entertained and engaged. Its goal as an opponent would not be to win but to provide optimal entertainment informed by an understanding the player's needs and capabilities.

Adult-Child Learning Games

Many games played between adults and children are focused on teaching children new skills or helping them with their cognitive development. These games could also be categorized as friendly games, but here we focus on the educational aspects. The relationships here might be parent/child or teacher/student. The analogy we make with superintelligence puts the AI in the role of adult teacher and the human in the role of child student.

Learning How to Play. Learning how to play is an important part of childhood. Following rules and understanding fairness takes practice, and learning how to develop strategies and skills to play games is an essential part of mental and physical development. *Candy Land* is a completely random board game with no strategy, but it teaches children how to follow instructions, play fair, and interact with game elements. *Tic-Tac-Toe* is a game that is trivial for adults to play perfectly, resulting in an inevitable draw, but which children still find entertaining. When playing this game against a child, an adult might choose to intentionally draw or lose instead of win, just to keep the child interested and learning. *Tee-Ball*, a simplified version of baseball where the ball starts stationary on a post in front of the batter, simplifies the rules of an adult game so that children can play it at their own skill level.

As game designers, we are especially interested in the idea of entry-level versions of games designed by superintelligence to be played between each other, but then "dumbed down" for humans to play. We also imagine superintelligences creating games that teach us how to better negotiate and find positive outcomes in complex conflict scenarios, perhaps using AI-generated models of human minds which we do not currently ourselves understand.

Learning Positive Behaviors. Some games are designed to teach morals and good behavior. Some versions of *Chutes & Ladders*, are designed to illustrate and reinforce obedience and etiquette. *Airplane Spoon Feeding* is a game played with very young children that can't yet feed themselves, to encourage them to eat when they might not want to. Adults also arrange games for children to keep them occupied or distract them from negative or destructive behavior.

A Superintelligence with deep models of human psychology could develop games that move us towards more healthier, happier lifestyles. To the degree that we can be certain that an advanced AI is fully committed to helping us achieve goals that actually correspond to our genuine (at times conflicting) values, we may welcome games that seek to persuade and manipulate us. But once we begin to submit ourselves to this kind of game we may lose the ability to distinguish between our genuinely held values and those our AI companions want us to have, at which point something important will have been lost. In general, there are significant challenges to resolving these kinds of transitional values conflicts (Brundage 2014).

Learning Knowledge Games like *Trivial Pursuit for Kids* can be educational in that they teach children facts, but this category can also include *Educational Games* offered by organizations such as BrainPOP or PBS which are intentionally focused on teaching math, English, history, or science. Making good educational games is a significant challenge, as what makes a good game and what makes a good lesson are not necessarily the same (Gee 2003). In addition, interesting subject matter for a game is not necessarily what we would like to teach.

We imagine a future where a general artificial intelligence understands and models human beings so well that they can create perfectly paced educational games to keep us engaged and learning at an optimal rate. They might be able to predict when we are about to be negatively frustrated and temporarily provide a slightly easier exercise or a break at just the right length of time. These games might be quite complicated indeed, teaching quantum mechanics and foreign languages.

Adult-Child Observation Games

Some games are played between children, where adults watch the children play without being an active participant. Examples of this type of play include parents and grandparents on a *Park Bench at a Playground* watching children playing happily, or *Child Sports Spectating* while children are playing organized sports. Adult-Child Observation Games also occur when developmental psychologists study how children play or learn; the purpose is for the psychologist to learn something about child behavior.

Superintelligences may watch and study us from afar for any number of reasons. They may be building more accurate models of the human mind, just to increase their own knowledge or so they can predict our actions better. They may experience the equivalent of aesthetic pleasure, or they may have reasons for wanting to observe our behavior that we are not even capable of comprehending.

Adult-Child Domination Games

Children do not always enjoy participating in the games that parents, coaches, or teachers force them to play. Perhaps the adult is *Forcing Play* by demanding that a child plays a game with them because the adult desires that the child become much better at playing the game (e.g. a parent wants their child to become a tennis professional). Certainly some children have spent a significant amount of time protesting and complaining about their overbearing parents who forced them to practice a game or skill they did not enjoy.

Domination games can also arise in the context of *Wrestling* or *Tickling* where an adult is physically playing with a child and can use their superior strength to overpower the child – at first this can be fun for the child, but quickly can go to far where the child is crying for the parent to stop.

An artificial intelligence may decide, against our will, that some games are good for us to play. As with the *Learning Positive Behaviors* section, the machines have made a value judgment for us, and we are no longer able to resist them. While in the long run, perhaps humanity gets some measurable benefit, but our experience may be one of suffering and crying in futility as we try to control our own destiny. And as in the example of *Tickling*, it can be difficult and perhaps even impossible to judge exactly when we have been pushed too far by the machines such that the activity goes from being pleasurable to upsetting.

Games Between Expert and Novice

For games that have a wide range of skill levels playing the same game, such as *Chess*, *Golf*, or *Go* we have conditions where the expert and novice players would like to play together, or the novice desires to learn from the expert.

Expert-Novice Friendly Games

Handicapping is the process of giving a lower skilled player an advantage so that the game is more balanced when playing against a better opponent. In sports like *Pick-up Basketball* the less skilled team can start with more points and the game is played until one team reaches a goal score. In *Go*, there is a formalized system of giving starting stones to the player with lower rank. *Golf* also has formalized handicapping where the less skilled player is given a scoring advantage.

When playing against a highly skilled AI in a strategic game, we certainly need a severe handicap to equalize its extreme advantages. We can also imagine that humans could be used as the handicap in team games between superintelligences. For example, two superintelligences of unequal computational power might have humans on their team to equalize the difference.

Expert-Novice Learning Games

Tutoring, *Mentoring*, and *Coaching* are examples where the expert player, either through volunteering or paid work, is teaching the novice how to play better through active teaching methods. The expert can get satisfaction at helping others get better at the game, and might even improve their own ability through explaining it to others. *Studying Losses* is the practice of intentionally playing against players that are much better than you without a handicap in order to improve and learn from one's errors.

Superintelligences map to the role of expert, teaching humans how to become better at games – in this mapping even our most brilliant Chess or Go experts would still be novices in comparison to the skill of a superintelligent AI, which could also have deep insight into human psychology and optimal training methods.

Expert-Novice Observation Games

Novices observe experts playing games in the domain of *Spectating Professional Sports*. Games focused on mental skills with well defined tournaments such as *Chess, Poker, and e-Sports Tournaments* are particularly relevant. We abuse the definition of "novice" here to refer to a player who is significantly less skilled, but not necessarily a beginner.

Novices enjoy watching the high quality of play that experts can achieve, giving us a glimpse at the remarkable abilities of great players. Some games are designed with spectating in mind – the professionals could play just for themselves, but the spectacle and athleticism they exhibit to their fans and audience is part of the enjoyment of the game. Novices may also observe experts in order to improve their own level of play.

It may be particularly interesting to observing the games that superintelligences might play with each other. These games will likely require a projection into a smaller mental space that humans can appreciate, accompanied by narration and commentary – either translated into a human understandable level by the superintelligences themselves or translated by humans attempting to understand what the superintelligences are doing (Chiang 2000).

Expert-Novice Domination Games

It is not surprising that an expert would win against a novice in an unevenly matched game, but here we are discussing expert-novice play where the superior player repeatedly beats players of obviously lesser skill for entertainment or rewards. In *Griefing* or *Player Killing*, a more advanced player intentionally harms other players in multiplayer games, most commonly in massively multiplayer online role playing games, and the "more massive the distress caused, the greater the killer's joy at having caused it" (Bartle 1996). In *Greed Play*, the expert player is trying to optimize their own game experience, perhaps trying to increase their gold or experience points by killing easy-to-beat novices, without caring about how this activity might affect the experience of other players (Foo and Koivisto 2004). This is also reflected in the shark/fish relationship in Poker, Pool, and other gambling games, where the dominating player extracts as much money from the lower skilled player as possible.

There is no guarantee that a superintelligent AI would care for our own feelings merely because it is intelligent (Bostrom 2014), so it might be beating us to optimize for its own utility functions without caring what auxiliary harm might be done to humans. A seed AI that was originally programmed to *win as many games as possible* might force all living humans to constantly play games against it: this might be the best way to win as often as possible. As with greed play, the machine isn't necessarily trying to punish us, but our welfare is less important to the superintelligence than its own.

Games Between Humans and Animals

The largest disparity in intelligence, and perhaps the most instructive, is when humans engage in games with animals. Human-Animal games can involve pets, such as dogs, cats, and hamsters, zoo animals such as horses, pigs, apes, and elephants, and/or wild animals like foxes, deer, and birds.

Although it can be uncomfortable to think of humans as being pets to a superintelligence, we believe it is a useful model to think about one possible way they may interact with us - protecting us and taking care of our physical and psychological needs, but also controlling and limiting our behavior. The degree to which this constitutes an existential danger is a larger topic we do not address in this paper.

For non-pet animals, there is a wide spectrum of how humans treat animals, extending from games that we play together with animals (i.e. polo and horse racing) to controversial sports that typical result in the death of the animal (i.e. hunting, fishing, or bull fighting). In particular, it would be an existential crisis if superintelligences were to play domination games with us.

Human-Animal Friendly Games

Friendly games between humans and animals are those where the animals are not harmed and the intention of the play is often to be entertaining and enjoyable for both orders of intelligence. We recognize the ethical issues surrounding some of these games and that the animals may indeed not be entertained by the game or the training for the game. But for our analogy of friendly games, we focus on games where the animals are not harmed during the game, as exemplified by research that provides positive playful experiences between humans and orangutans (Wirman 2014) or humans and pigs (Driessen et al. 2014).

Providing a Playful Environment. We can provide an intellectually stimulating environment to our pets, such as providing hamster trails or hamster wheels, scratching posts and climbing posts for cats, or providing toys for dogs to chew on. This allows the animal to play and explore on its own, although we are providing the materials which define the play experience. With the lowest order of intelligence, such as reptiles and fish, much of the "play" comes from providing them food. Giving a pet something new to eat, or something that must be engaged with to eat, can be abstracted as an intellectually stimulating or playful experience.

A superintelligence might provide us with new and interesting games to play, to fulfill our needs to engage in playful activities. An artificial intelligence might study a group of humans, design a game for the target audience, watch them play and continuously update it to keep it stimulating and entertaining. Researchers working on procedural game generation are already engaged in making game systems that can create novel games (Togelius and Schmidhuber 2008; Browne and Maire 2010; Cook and Colton 2011; Treanor et al. 2012; Isaksen et al. 2015).

A superintelligence might also play matchmaker, and decide which humans might like to play together either cooperatively or competitively. Matching algorithms already exist for multiplayer games, but the example here would be for significantly higher levels of predictive accuracy.

Pure Play. We play directly with our pets, such as tossing a ball and playing fetch with a dog, or shaking a string or creating a dancing laser dot for a cat. In addition to providing

exercise for the pet, it provides us joy to see our pets having fun. It also helps us model the cognitive abilities of the pet, such as watching they them try to find us when hiding behind a door or curtain. It's a source of enjoyment to observe the unpredictability that can occur when watching a pet chase a ball repeatedly or try to get two large chew toys in its mouth at the same time, coming up with different strategies and methods to complete its goal.

A superintelligent form of pure play might be *Practical Jokes* played on humans, e.g. by interfering with a social situation and watching the outcome. The challenge here for the superintelligence is to make the joke funny, not cruel, for human participants.

Human-Animal Team Play. *Polo* is the prime example of humans and horses working together in a team, celebrating the abilities that each species excels at. Although humans could play a game like Polo without horses, its the combination that is a key characteristic of the game. *Equestrian* (horseback riding at the Olympics), *Vaulting* (gymnastics performed by humans while on horseback), *Horse Racing*, *Chariot Racing*, and *Dogsled Racing* are other games that are played with humans and animal teams. These animals are well trained and work closely with humans to perform at their best as a team.

Something similar happens with *Centaur Chess*, where humans work together with a computer chess algorithm to compete against other human+computer teams. This celebrates the combination of superior computer tactical planning with human creativity and strategy, performing better than chess computers alone (Brynjolfsson and McAfee 2012).

Games designed for superintelligent and human hybrid teams are exciting for us to think about. Although superintelligences could play more perfectly without humans on their teams, the human element adds drama and uncertainty that would not be available to strategically optimal machines. Imagine, for example, a game of complex resource management with alliances and political negotiation, played by 100 humans each with an advanced AI companion whispering advice in their ear. The AIs' ability to navigate the complex problem-space will be refracted through the essentially human qualities of charisma, persuasion, duplicity, and trust.

Human-Animal Learning Games

We teach *Dog Tricks*, such as teaching our dogs to sit, shake, or high-five, which provides mental stimulation for both the owner and pet as we learn how to communicate our desires to the dog in a language they can understand. These tricks provide entertainment to our friends and guests, and demonstrate that the pet is reliable, well-trained, and well-behaved. This also helps us better understand and model the limits of the animal's intelligence.

We have discussed learning in the Adult/Child section, but those games are focused on teaching skills that children will need when they grow up to be well-functioning adults. Here we think about games that will teach topics that might not be economically valuable in a society featuring superintelligences, but that humans would like to learn anyway merely because they are interested in them.

Related to how we show off dog tricks to our friends, perhaps superintelligences might compete with each other regarding how advanced the skills they can teach humans, communicating "Look what my human can do!"

Human-Animal Observation Games

Humans enjoy watching animals in a natural or synthetic habitats. In *Birdwatching*, people enjoy tracking, cataloging, and photographing birds found in the wild. For synthetic habitats, it can be relaxing, entertaining, and educational to watch fish in a *Fish Tank* gently swimming, ants building tunnels in an *Ant Farm*, or monkeys swinging on branches in a *Zoo Exhibit*. In these cases, humans do not interact directly with the animals, but we extract enjoyment nonetheless.

We also enjoy watching other orders of intelligence play with each other, such as dogs at a *Dog Park* (and from the section on Adult/Child games, this would apply to parents watching children play with their friends at a playground). Pets can exhibit neotenic behaviors and traits which we normally ascribe to childlike behavior in humans. This can remind us of our own childlike experiences, providing us with a bridge to another time in our lives. We also observe play to learn more about the animal psychology of the animals engaged in play.

This ties in closely with the Adult-Child Observation Games presented earlier, where the superintelligences learn about human behavior and improve their models. The idea of observation games is also linked with the idea of playful environments: just as humans have setup the conditions for fish tanks, ant farms, and zoo exhibits, a superintelligence might choose to setup fictional or historical environments for us to explore.

Human-Animal Domination Games

Humans play sports and games which focus on the domination of animals; it would be an existential crisis if superintelligences were to play these types of games with us. This category of games should be the most alarming to game and AI researchers, and provide us with a reason for taking this subject seriously. We present this catalog of domination human-animal games to make the analogy that not all games between superintelligences and humans are desirable.

Animal vs Animal. *Cock Fights*, *Dog Fights*, and *Dog Races* are examples of games where animals are pitted against each other in direct competition, often including violent, even deadly, combat. These games are often associated with gambling, and, despite their brutality, can have complex cultural meanings (Geertz 2005). These games have prompted serious ethical debate and are illegal in most contemporary societies.

We can imagine superintelligent machines gambling on human vs human competition, thought not necessarily in violent combat, as it could be in the domain of sports or other competitions.

Human vs Animal. *Hunting* and *Fishing*, when engaged in purely for entertainment, are examples of Human vs Animal games. Some of these activities offer the chance at danger or death for the human, such as *Bull Fighting* or *Running of the*

Bulls. These types of games don't always need to lead to the death of the animal, e.g. *Rodeo*. This category also includes the cruel practice of torture-games that children inflict on lower animals, like burning ants or pouring salt on snails.

The idea of superintelligences harming humans for sport is obviously horrifying. Why would a superintelligence be motivated by cruelty and domination? Perhaps this type of sport could arise from a whole-brain emulation where the human brain is algorithmically replicated by a machine as the base for a seed AI. In this case, if there is a hardwired human disposition to dominate less intelligent lifeforms, that behavior would be inherited by superintelligent AIs and inflicted on us. Another possibility is that humans own or control resources which are desired by the artificial intelligence; this is analogous to eating an animal killed while hunting or fishing.

Breeding and Dog Shows. *Dog Shows* are competitive activities with well defined rules designed to evaluate humans' ability to breed and train dogs (American Kennel Club 2015).

It is possible, if highly disturbing, to imagine a superintelligent analogue in which AIs compete to breed and train humans according to various criteria. The ethics of human gene selection are related to this discussion (Harris 1992).

Biotic Games. Researchers have invented *Biotic Games* that are played using live bacteria as game elements (Riedel-Kruse et al. 2011). The ethics of using bacteria for entertainment has been addressed, with the claim that we eliminate bacteria constantly (though not for entertainment), there are limitless numbers of them, and they are not self-aware.

A superintelligence which has orders of magnitude more senses and complexity could consider us too trivial or numerous to worry about, but we can also imagine a slightly more positive scenario where we are used as game pieces between two or more superintelligences, but are not disposed of after each game. This still keeps us in the realm of domination games, but is more closely related to dog races since the dogs are not intentionally harmed as part of the race itself.

Mixed Games. *Fox Hunting* is a complicated, mixed example, as humans are engaging in a sport leading to the death of a fox, but dogs and horses are participating as hunting partners. The relationship of hunter to dog is not a dominating one, but a friendly one, as they are working together.

This relates to the team based games we described earlier. Perhaps there are games in which superintelligences would take on the role of hunter and humans the role of dog. Of course, these mixed team games don't need to be hunting related; superintelligences and humans could work together towards any common game goal.

Conclusion

We have presented a survey of how a superintelligence might one day play games with humans, using analogies of how humans play games with different orders of skill and intelligence, namely Adult-Child play, Expert-Novice play, and Human-Animal play. Although we do not claim to know the motivations that superintelligences might have, we can at least use the analogies to guess at what types of game experiences may come as our lives become even more integrated with artificial intelligence.

It's particularly important to be aware of the domination games, as a warning to the procedural game generation and game AI communities that some type of morality checks and ethical guidelines will someday be necessary to integrate into our systems. Simply inheriting existing human ethics will not be sufficient, as shown by analogy with how humans currently play domination games against children, novices, and animals.

We plan to revisit these issues related to superintelligence and games, exploring exactly what it means for logic-based machines to play, to deeper examine how this research can improve existing game design, and how to encode some of these ethical issues into content generation systems.

Acknowledgments

We'd like to thank Barry Bunin, Rob Dubbin, Zach Gage, Kevin Cancienne, Ron Carmel, Dan Gopstein and Galen Guengerich for their insightful discussions about games and superintelligence, and for the anonymous reviewers for their comments and suggestions. We also would like to acknowledge Trixie for providing inspiration for games that people play with pets.

References

Alcock, J. 1993. *Animal behavior: An evolutionary approach*. Sinauer Associates.

American Kennel Club. 2015. *Rules Applying to Dog Shows*.

Barrat, J. 2013. *Our final invention: Artificial intelligence and the end of the human era*. Macmillan.

Bartle, R. 1996. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research* 1(1):19.

Bostrom, N. 2006. What is a singleton. *Linguistic and Philosophical Investigations* 5(2):48–54.

Bostrom, N. 2014. *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Browne, C., and Maire, F. 2010. Evolutionary game design. *Computational Intelligence and AI in Games, IEEE Transactions on* 2(1):1–16.

Brundage, M. 2014. Limitations and risks of machine ethics. *Journal of Experimental & Theoretical Artificial Intelligence* 26(3):355–372.

Brynjolfsson, E., and McAfee, A. 2012. *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee.

Bryson, J. J.; Kime, P. P.; and Zürich, C. 2011. Just an artifact: why machines are perceived as moral agents. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, 1641.

Caillois, R., and Barash, M. 1961. *Man, play, and games*. University of Illinois Press.

Chiang, T. 2000. Catching crumbs from the table. *Nature* 405(6786).

Cook, M., and Colton, S. 2011. Multi-faceted evolution of simple arcade games. In *CIG*, 289–296.

Driessen, C.; Alfrink, K.; Copier, M.; Lagerweij, H.; and van Per, I. 2014. What could playing with pigs do to us. *Game design as multispecies philosophy. Antennae: The Journal of Nature in Visual Culture* 30:79–102.

Foo, C. Y., and Koivisto, E. M. 2004. Defining grief play in mmorpgs: player and developer perceptions. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 245–250. ACM.

Fox, J., and Shulman, C. 2010. Superintelligence does not imply benevolence. *ECAP* 10:456–462.

Gee, J. P. 2003. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)* 1(1):20–20.

Geertz, C. 2005. Deep play: Notes on the balinese cockfight. *Daedalus* 134(4):56–86.

Genesereth, M.; Love, N.; and Pell, B. 2005. General game playing: Overview of the aaai competition. *AI magazine* 26(2):62.

Goertzel, B., and Pennachin, C. 2007. *Artificial general intelligence*, volume 2. Springer.

Good, I. J. 1965. Speculations concerning the first ultraintelligent machine. *Advances in computers* 6(99):31–83.

Graeber, D. 2014. What's the point if we can't have fun? *The Baffler*.

Harris, J. 1992. Wonderwoman and superman: the ethics of human biotechnology.

Huizinga, J. 1949. *Homo Ludens*. Taylor & Francis.

Isaksen, A.; Gopstein, D.; Togelius, J.; and Nealen, A. 2015. Discovering unique game variants. *Computational Creativity and Games Workshop at the 2015 International Conference on Computational Creativity*.

Koster, R. 2005. *Theory of fun for game design*. O'Reilly Media, Inc.

Kurzweil, R. 2005. *The singularity is near: When humans transcend biology*. Penguin.

Mateas, M., and Sengers, P. 2003. *Narrative intelligence*, volume 46. John Benjamins Publishing.

Newborn, M. 1997. *Kasparov Vs. Deep Blue: Computer Chess Comes of Age*. Springer.

Perez, D.; Samothrakis, S.; Togelius, J.; Schaul, T.; Lucas, S.; Couëtoux, A.; Lee, J.; Lim, C.-U.; and Thompson, T. 2015. The 2014 General Video Game Playing Competition. *IEEE Transactions on Computational Intelligence and AI in Games* (to appear) DOI: 10.1109/TCIAIG.2015.2402393.

Prensky, M. 2005. Computer games and learning: Digital game-based learning. *Handbook of computer game studies* 18:97–122.

Riedel-Kruse, I. H.; Chung, A. M.; Dura, B.; Hamilton, A. L.; and Lee, B. C. 2011. Design, engineering and utility of biotic games. *Lab on a Chip* 11(1):14–22.

Sandberg, A., and Bostrom, N. 2008. Whole brain emulation.

Schaeffer, J.; Burch, N.; Björnsson, Y.; Kishimoto, A.; Müller, M.; Lake, R.; Lu, P.; and Sutphen, S. 2007. Checkers is solved. *Science* 317(5844):1518–1522.

Schaul, T.; Togelius, J.; and Schmidhuber, J. 2011. Measuring intelligence through games. *arXiv preprint arXiv:1109.1314*.

Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* 2(3):230–247.

Sutton-Smith, B. 2009. *The ambiguity of play*. Harvard University Press.

Togelius, J., and Schmidhuber, J. 2008. An experiment in automatic game design. In *Computational Intelligence and Games, 2008. CIG'08. IEEE Symposium On*, 111–118. IEEE.

Togelius, J.; Yannakakis, G. N.; Stanley, K. O.; and Browne, C. 2011. Search-based procedural content generation: A taxonomy and survey. *Computational Intelligence and AI in Games, IEEE Transactions on* 3(3):172–186.

Togelius, J.; Shaker, N.; and Yannakakis, G. N. 2014. Active player modelling. In *Proceedings of Foundations of Digital Games*.

Treanor, M.; Blackford, B.; Mateas, M.; and Bogost, I. 2012. Game-o-matic: Generating videogames that represent ideas. In *Proceedings of the The third workshop on Procedural Content Generation in Games*, 11. ACM.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 433–460.

Vinge, V. 1993. The coming technological singularity: How to survive in the post-human era. *Vision-21: Interdisciplinary science and engineering in the era of cyberspace* 11–22.

Vygotsky, L. S. 1967. Play and its role in the mental development of the child. *Soviet psychology* 5(3):6–18.

Wirman, H. 2014. Games for/with strangers-captive orangutan (pongo pygmaeus) touch screen play. *Antennae* (30):105–115.

Young, R. M.; Riedl, M. O.; Branly, M.; Jhala, A.; Martin, R.; and Saretto, C. 2004. An architecture for integrating plan-based behavior generation with interactive game environments. *Journal of Game Development* 1(1):51–70.

Yudkowsky, E. 2001. Creating friendly ai 1.0 – the analysis and design of benevolent goal architectures.

Zimmerman, E., and Chaplin, H. 2013. Manifesto: The 21st century will be defined by games. http://kotaku.com/manifesto-the-21st-century-will-be-defined-by-games-1275355204.