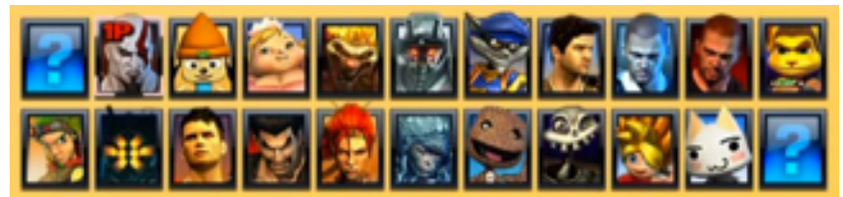
Today I'll be doing a deep dive into one aspect of game balance. It's the 25-minute lightning talk version of the 800-minute ring cycle talk. I'll be developing one big idea, so hold on tight!

In the fall of 2012, I worked as a technical designer, using data to help balance PlayStation All-Stars Battle Royale, Superbot and Sony Santa Monica's four-player brawler featuring PS characters from throughout the years. Btw, the game is great, and surprisingly unique, despite its well-known resemblance to that other four-player mascot brawler, Shrek Super Slam. My job was to make sense of the telemetry we gathered and use it to help the designers balance every aspect of the game. What were the impacts of skill, stage, move selection, etc. and what could that tell us about what the game feels like to play?

Balance

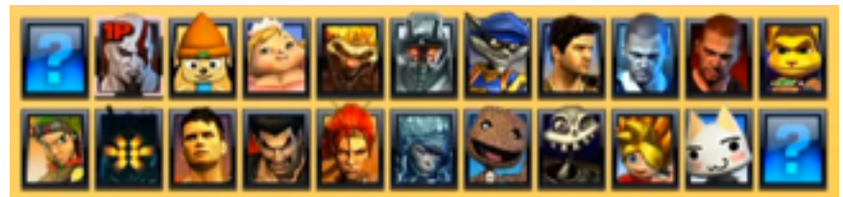
In the fall of 2012, I worked as a technical designer, using data to help balance PlayStation All-Stars Battle Royale, Superbot and Sony Santa Monica's four-player brawler featuring PS characters from throughout the years. Btw, the game is great, and surprisingly unique, despite its well-known resemblance to that other four-player mascot brawler, Shrek Super Slam. My job was to make sense of the telemetry we gathered and use it to help the designers balance every aspect of the game. What were the impacts of skill, stage, move selection, etc. and what could that tell us about what the game feels like to play?

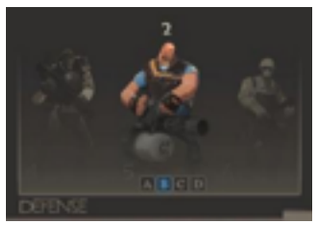
Character Balance



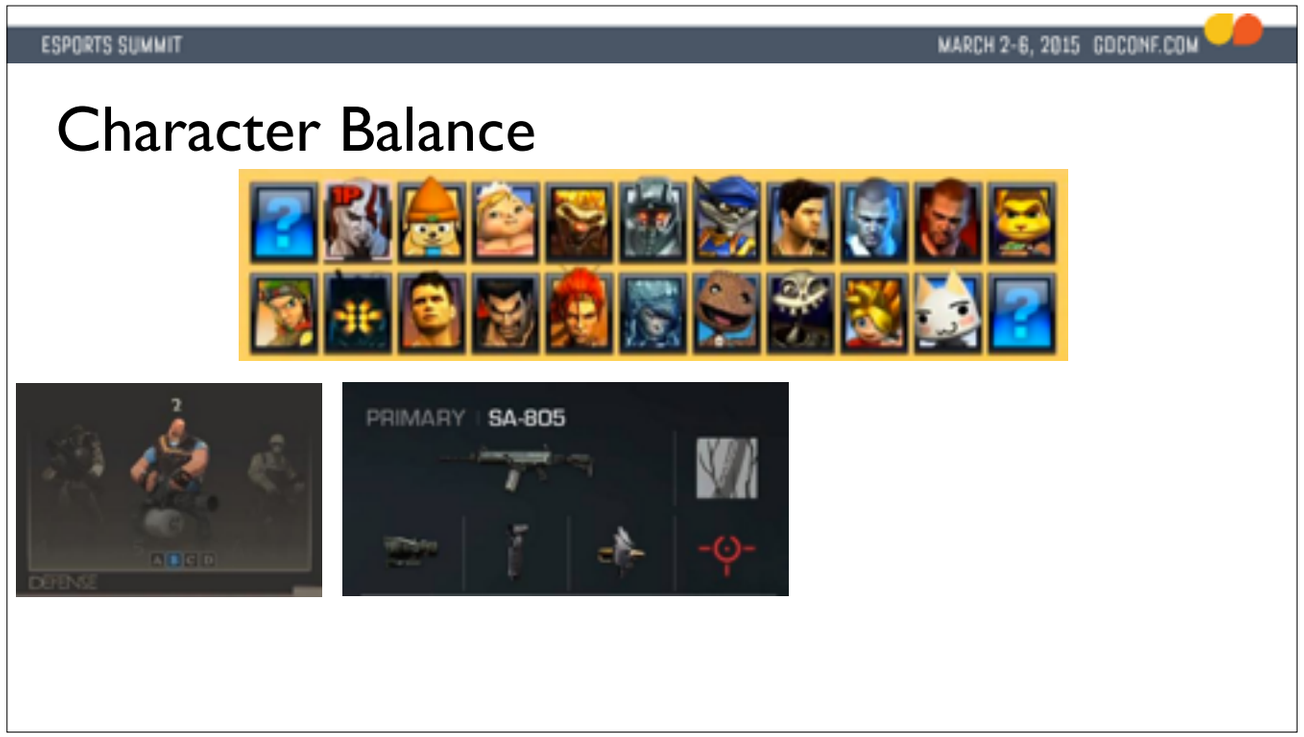
There were a lot of big questions about what it's like to play the game, but at least one pretty quantitative question was paramount: are these characters balanced with respect to one another? I.e. is any character underpowered or overpowered? Are there many viable asymmetric options? I'm essentially talking about what the metagame itself feels like to play. Do I have freedom, and meaningful options, in the metagame. Can I play the character I love and not feel cheated? This question isn't unique to fighting games - most eSports wrestle with it. Whether we're talking about classes in TF2, or loadout archetypes in CoD, or races in Starcraft, or team dynamics in LoL, we take seriously the mission of trying to avoid seriously overpowered or underpowered asymmetric options. Players have more options, the game feels more fair, spectators and players both get more variety. Note that the ideas in this talk are relevant to all these games - with some modification. I'll be focusing on fighting games since they have one of the simplest metagames.

Character Balance

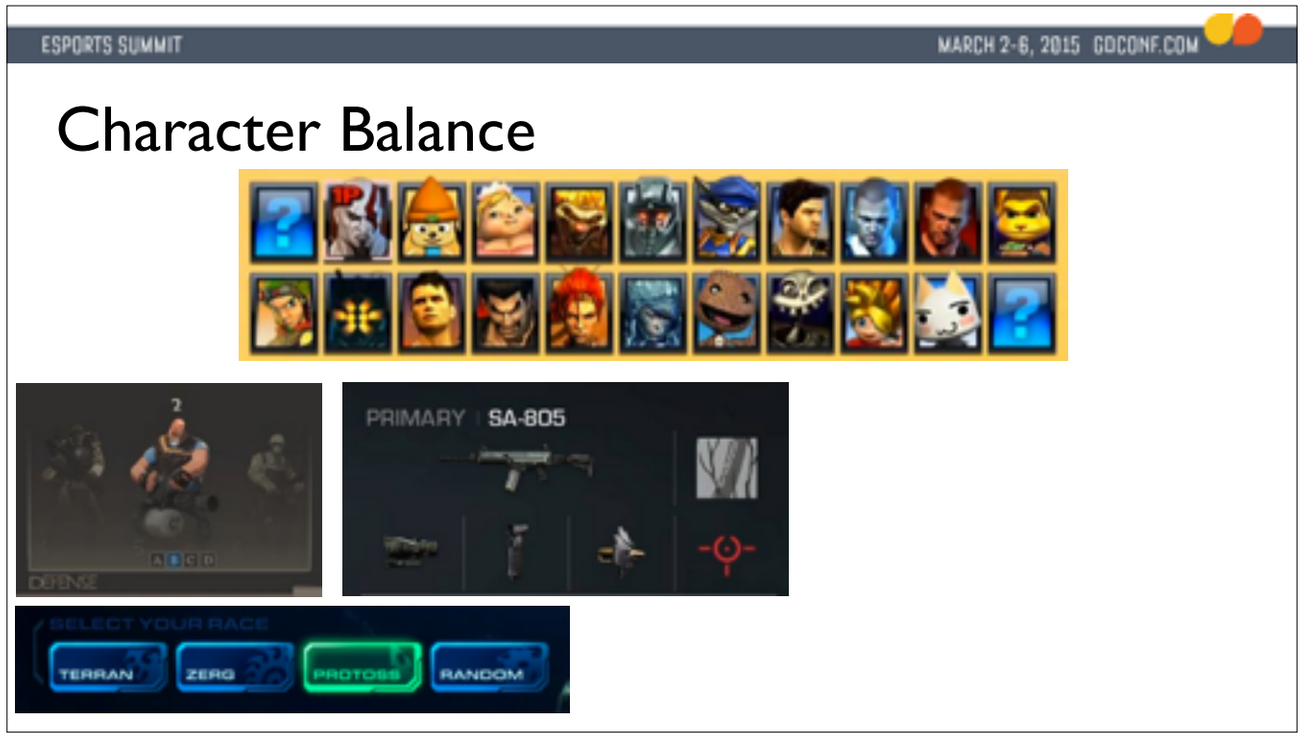




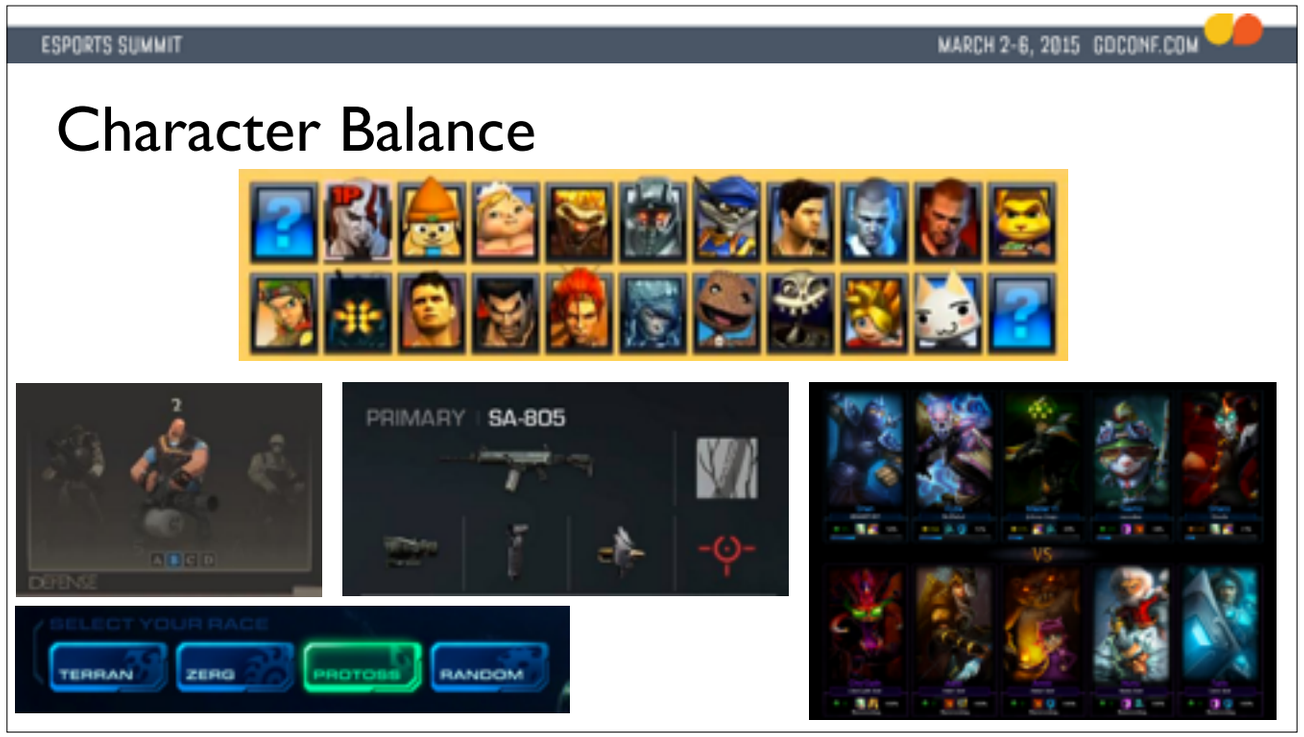
There were a lot of big questions about what it's like to play the game, but at least one pretty quantitative question was paramount: are these characters balanced with respect to one another? I.e. is any character underpowered or overpowered? Are there many viable asymmetric options? I'm essentially talking about what the metagame itself feels like to play. Do I have freedom, and meaningful options, in the metagame. Can I play the character I love and not feel cheated? This question isn't unique to fighting games - most eSports wrestle with it. Whether we're talking about classes in TF2, or loadout archetypes in CoD, or races in Starcraft, or team dynamics in LoL, we take seriously the mission of trying to avoid seriously overpowered or underpowered asymmetric options. Players have more options, the game feels more fair, spectators and players both get more variety. Note that the ideas in this talk are relevant to all these games - with some modification. I'll be focusing on fighting games since they have one of the simplest metagames.



There were a lot of big questions about what it's like to play the game, but at least one pretty quantitative question was paramount: are these characters balanced with respect to one another? I.e. is any character underpowered or overpowered? Are there many viable asymmetric options? I'm essentially talking about what the metagame itself feels like to play. Do I have freedom, and meaningful options, in the metagame. Can I play the character I love and not feel cheated? This question isn't unique to fighting games - most eSports wrestle with it. Whether we're talking about classes in TF2, or loadout archetypes in CoD, or races in Starcraft, or team dynamics in LoL, we take seriously the mission of trying to avoid seriously overpowered or underpowered asymmetric options. Players have more options, the game feels more fair, spectators and players both get more variety. Note that the ideas in this talk are relevant to all these games - with some modification. I'll be focusing on fighting games since they have one of the simplest metagames.



There were a lot of big questions about what it's like to play the game, but at least one pretty quantitative question was paramount: are these characters balanced with respect to one another? I.e. is any character underpowered or overpowered? Are there many viable asymmetric options? I'm essentially talking about what the metagame itself feels like to play. Do I have freedom, and meaningful options, in the metagame. Can I play the character I love and not feel cheated? This question isn't unique to fighting games - most eSports wrestle with it. Whether we're talking about classes in TF2, or loadout archetypes in CoD, or races in Starcraft, or team dynamics in LoL, we take seriously the mission of trying to avoid seriously overpowered or underpowered asymmetric options. Players have more options, the game feels more fair, spectators and players both get more variety. Note that the ideas in this talk are relevant to all these games - with some modification. I'll be focusing on fighting games since they have one of the simplest metagames.



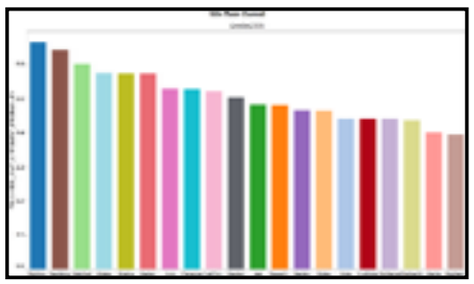
There were a lot of big questions about what it's like to play the game, but at least one pretty quantitative question was paramount: are these characters balanced with respect to one another? I.e. is any character underpowered or overpowered? Are there many viable asymmetric options? I'm essentially talking about what the metagame itself feels like to play. Do I have freedom, and meaningful options, in the metagame. Can I play the character I love and not feel cheated? This question isn't unique to fighting games - most eSports wrestle with it. Whether we're talking about classes in TF2, or loadout archetypes in CoD, or races in Starcraft, or team dynamics in LoL, we take seriously the mission of trying to avoid seriously overpowered or underpowered asymmetric options. Players have more options, the game feels more fair, spectators and players both get more variety. Note that the ideas in this talk are relevant to all these games - with some modification. I'll be focusing on fighting games since they have one of the simplest metagames.

Evaluating Character Strength

Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

Evaluating Character Strength

Win Rate

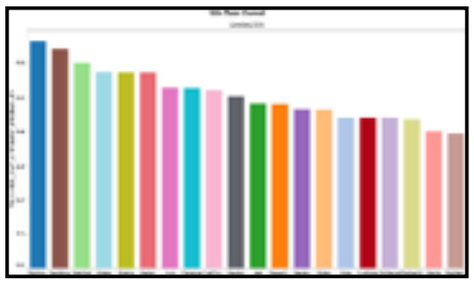


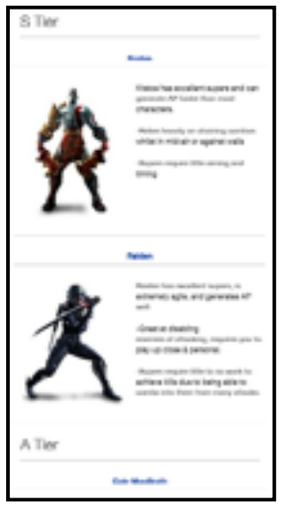
Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

Evaluating Character Strength

Tier Lists

Win Rate



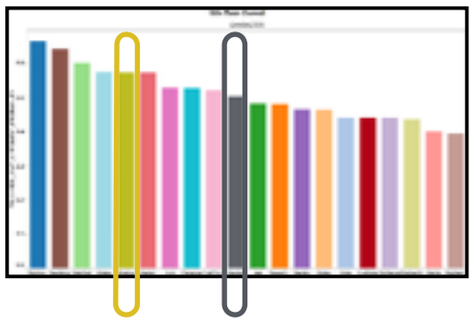


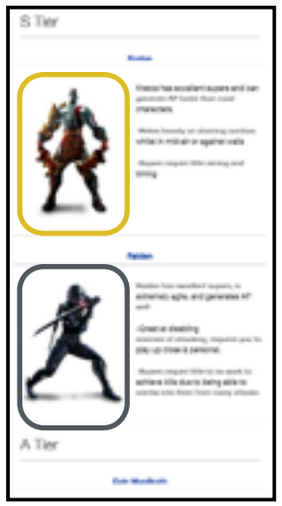
Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

Evaluating Character Strength

Tier Lists

Win Rate



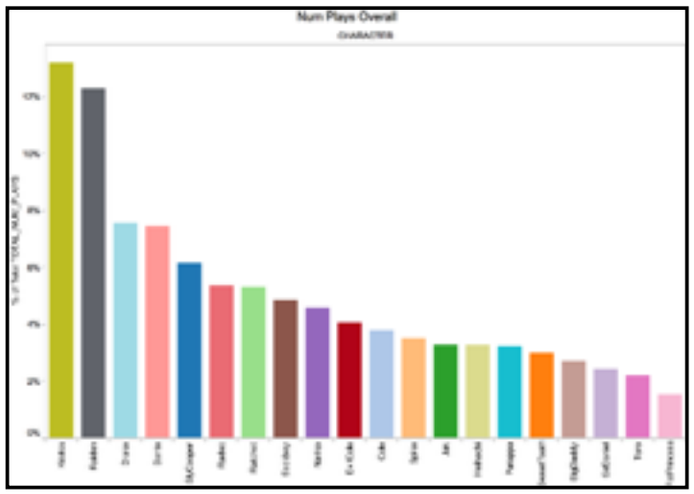


Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

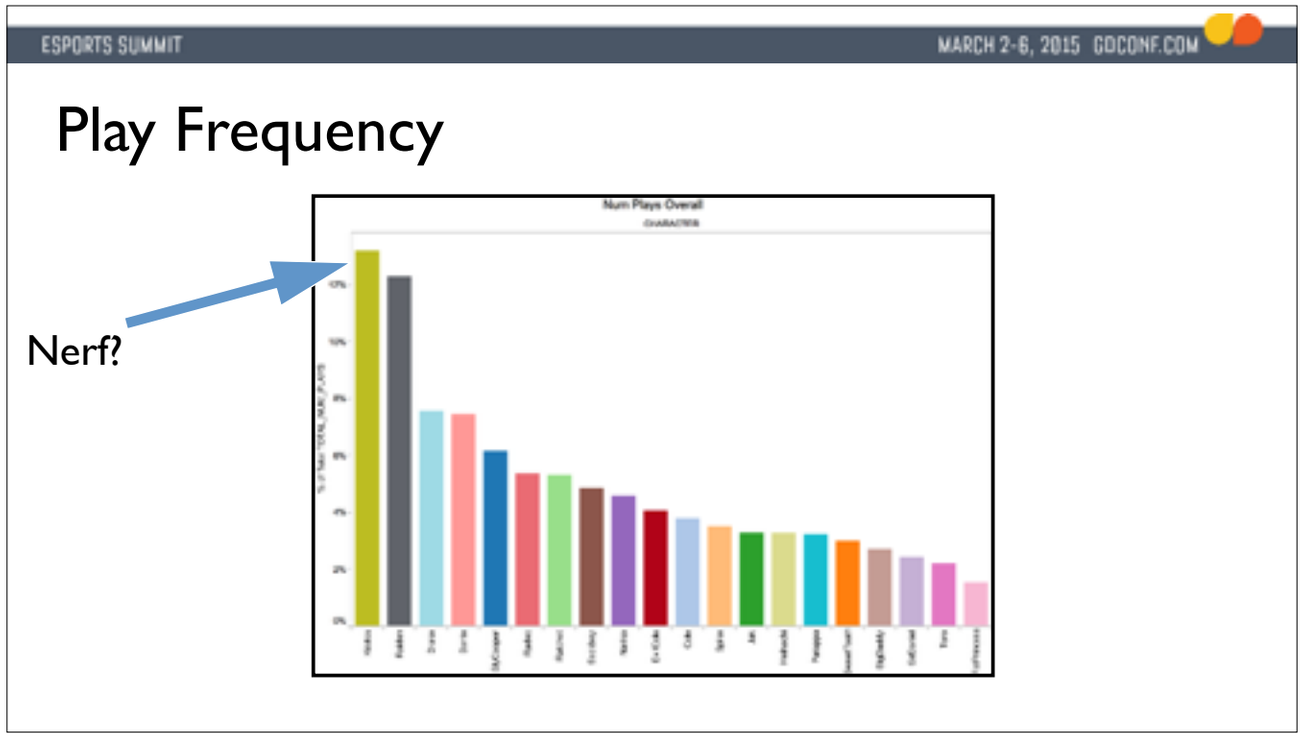
Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

Evaluating character strength is more easily said than done. It's not a well-defined concept; designers and players have numerous seemingly conflicting notions of strength. I want to focus on two. If we're collecting data, we can look at win rates, i.e. the fraction of games they win (perhaps restricted to high skill players). Or if we're feeling masochistic, we can listen to the players, as in this IGN tier list. Not masochistic because they're wrong necessarily, but because it calls into question your entire notion of who's strong and who's weak. The tier lists we saw had very little correlation with empirical win rates, at any skill. Kratos and Raiden dominated most of these tier lists, yet showed up as middle and upper-middle in the win rates. It's tempting to brush off players' concerns, and say that players don't know what they're talking about, or just like dudes with big swords. But our job is to create a good experience, and there was clear frustration. Indeed when we went and patched the game, we tried to address both win rates and player-perceived tier lists. I couldn't help but feel like there might be some unifying notion of character strength that united these perspectives - some rigorous reason why tier lists can and should be considered. And indeed, although tier lists didn't correlate with win rates, they did correlate with something else: play frequency. Those huge bars at the left are players' two "top tier" characters, Kratos and Raiden. 13% of all players were playing Kratos, which meant, in a 4-player game, that 43% of games had at least one Kratos. No wonder players were frustrated. A large fraction of their games - and consequentially their losses - were to these two characters.

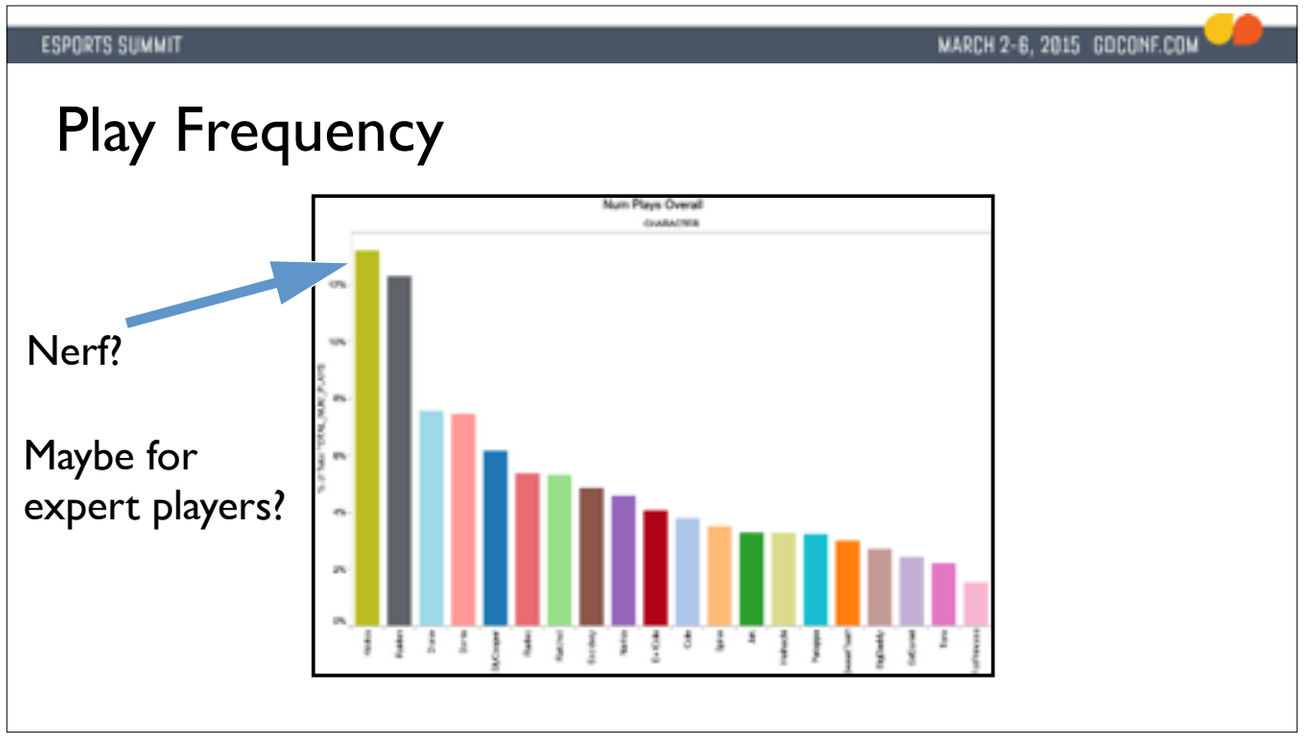
Play Frequency



So now let's entertain a naive idea. Can we just treat high play rates as a sign of overpowered characters? In other words, nerf (i.e. weaken) the frequently played characters? Obviously a lot of personal preferences play into players' choice of characters, like prior fan bases. But lets take for granted that expert players make their choice of character primarily "rationally", that is, to win. Then play rates at the expert level should implicitly act as a signal of win rate anyway. Expert players mostly play the characters who win.



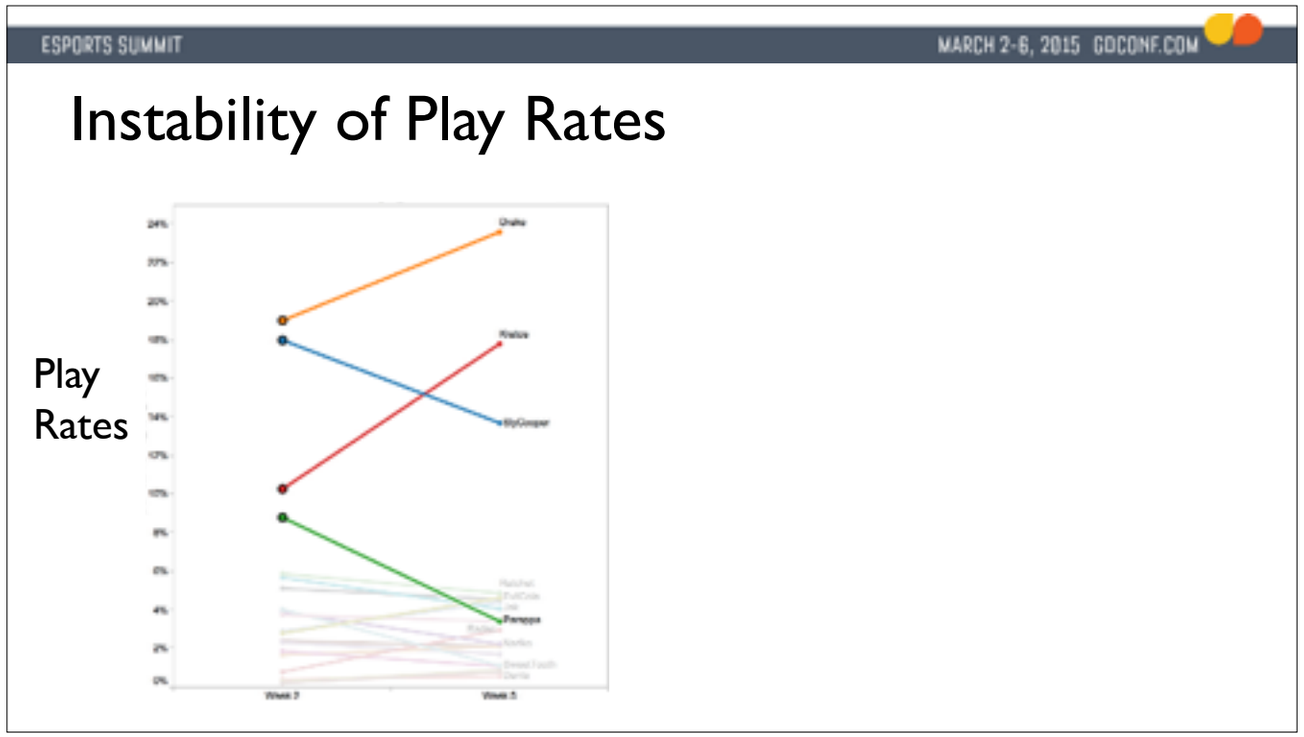
So now let's entertain a naive idea. Can we just treat high play rates as a sign of overpowered characters? In other words, nerf (i.e. weaken) the frequently played characters? Obviously a lot of personal preferences play into players' choice of characters, like prior fan bases. But lets take for granted that expert players make their choice of character primarily "rationally", that is, to win. Then play rates at the expert level should implicitly act as a signal of win rate anyway. Expert players mostly play the characters who win.



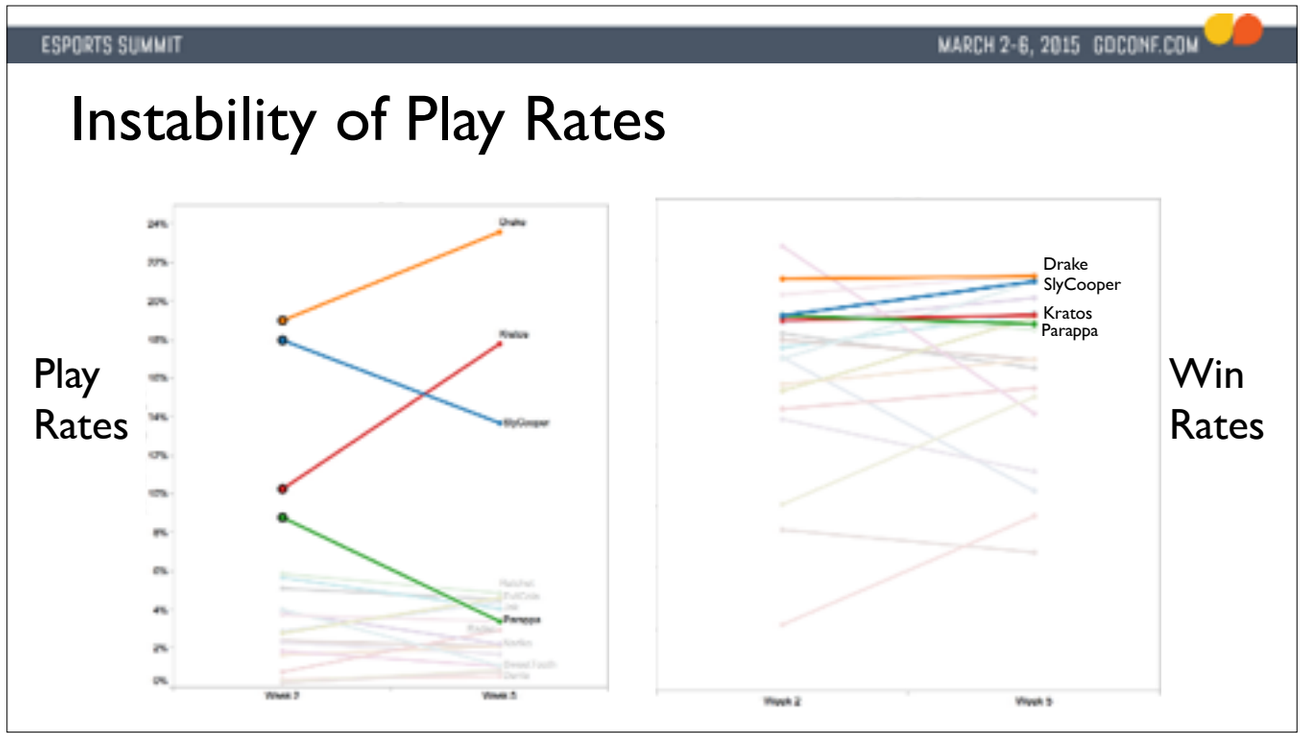
So now let's entertain a naive idea. Can we just treat high play rates as a sign of overpowered characters? In other words, nerf (i.e. weaken) the frequently played characters? Obviously a lot of personal preferences play into players' choice of characters, like prior fan bases. But lets take for granted that expert players make their choice of character primarily "rationally", that is, to win. Then play rates at the expert level should implicitly act as a signal of win rate anyway. Expert players mostly play the characters who win.

Instability of Play Rates

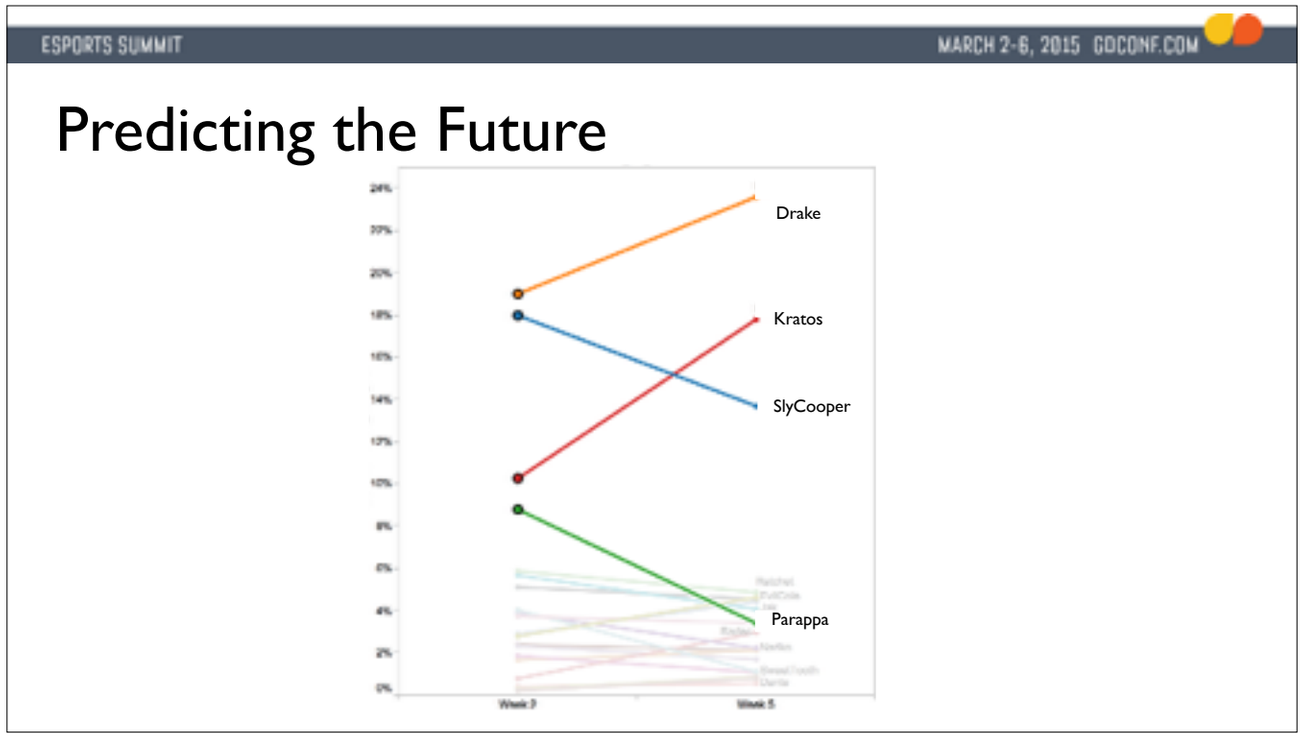
There's a problem though. Even if you look at play rates of top players, they're very unstable. Here's a graph of expert play rates after our first patch, in the second week and the fifth week. Four characters are highlighted. These numbers are all over the place! I said that expert players pick their characters to win, so why are they moving around so much? Maybe players are figuring out new ways to use these characters post-patch, and they're getting more popular as they win more. But lo! Here are the win rates for the same time period. This graph is pretty zoomed in, too. Almost no variation. So what's going on? Why are play rates changing so much? Well one problem is that it takes players a couple weeks to all gather around an overpowered option, even after its discovered. But the bigger problem, as usual, is free will. In a balanced game there are a lot of ways players can choose characters to win! What's more, these choices exert pressure on each other, so as Kratos gets more popular, Parappa - who has a disadvantage against him - is forced to play less. This creates a potentially never-ending churn of trends in the metagame, even without changes to the game or tactics.



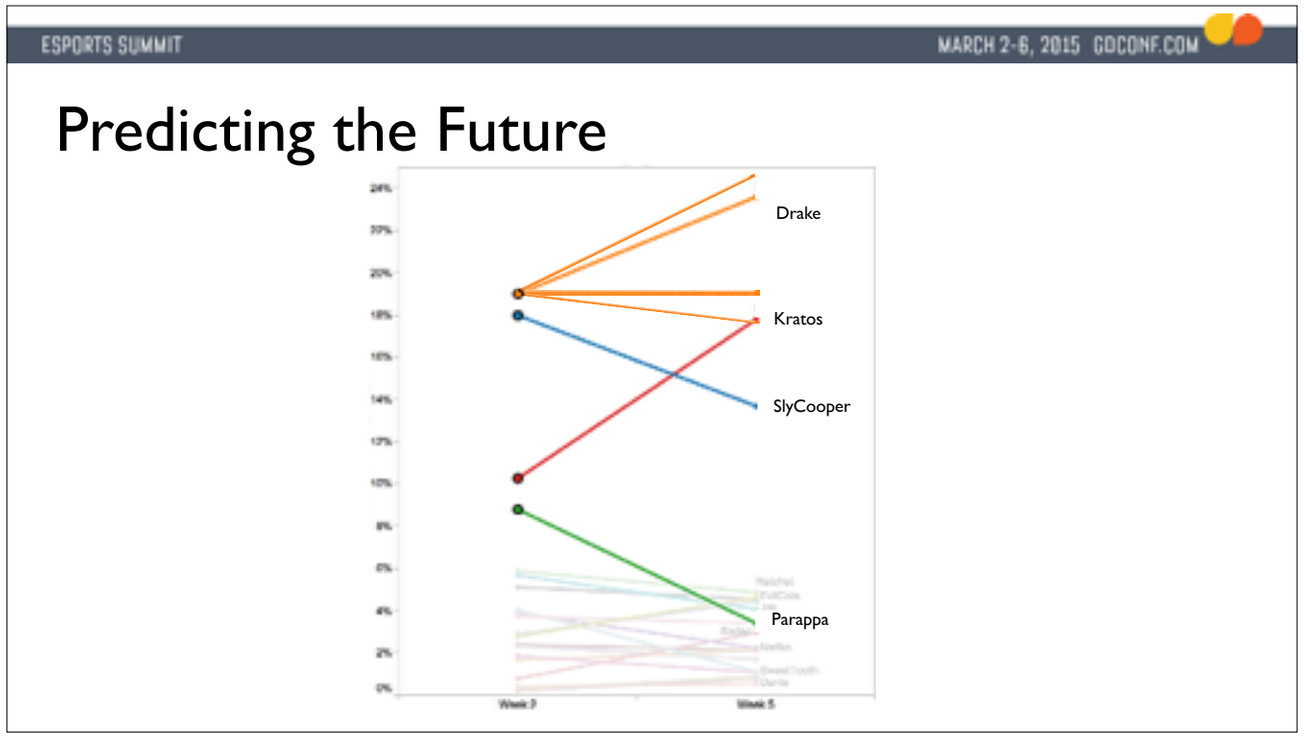
There's a problem though. Even if you look at play rates of top players, they're very unstable. Here's a graph of expert play rates after our first patch, in the second week and the fifth week. Four characters are highlighted. These numbers are all over the place! I said that expert players pick their characters to win, so why are they moving around so much? Maybe players are figuring out new ways to use these characters post-patch, and they're getting more popular as they win more. But lo! Here are the win rates for the same time period. This graph is pretty zoomed in, too. Almost no variation. So what's going on? Why are play rates changing so much? Well one problem is that it takes players a couple weeks to all gather around an overpowered option, even after its discovered. But the bigger problem, as usual, is free will. In a balanced game there are a lot of ways players can choose characters to win! What's more, these choices exert pressure on each other, so as Kratos gets more popular, Parappa - who has a disadvantage against him - is forced to play less. This creates a potentially never-ending churn of trends in the metagame, even without changes to the game or tactics.



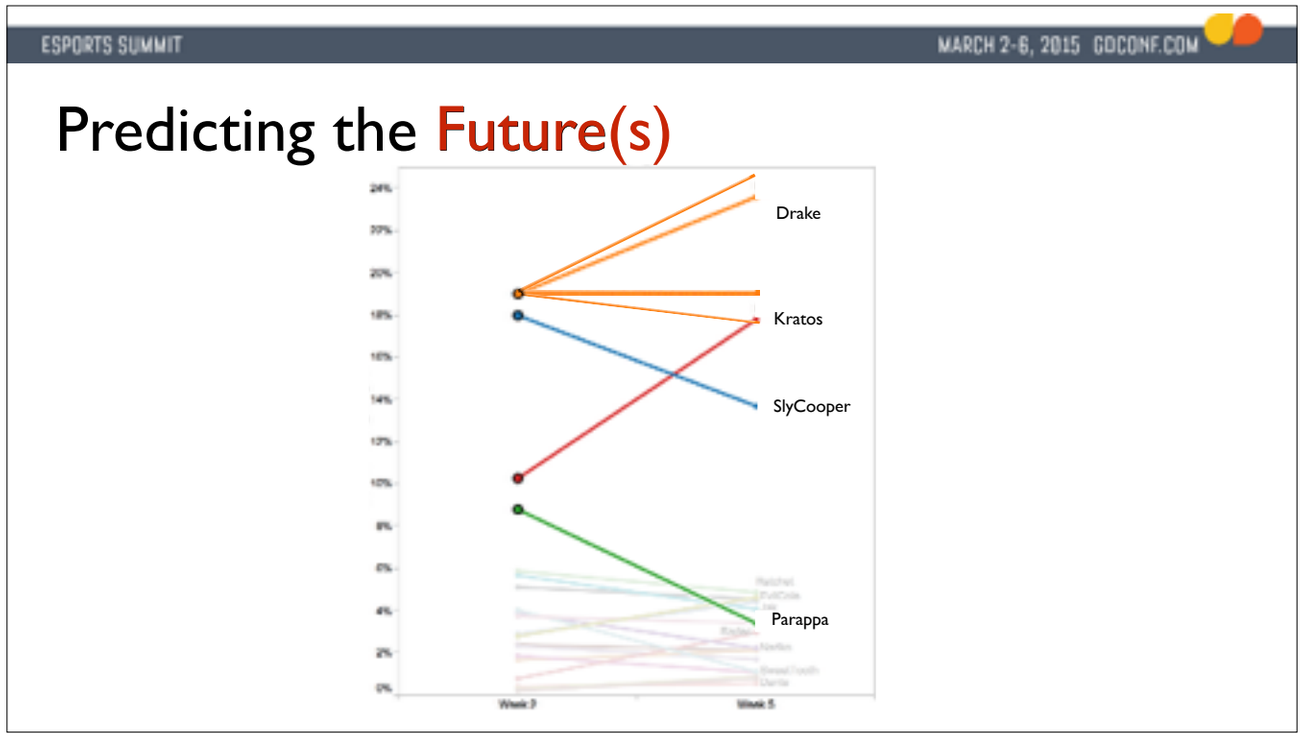
There's a problem though. Even if you look at play rates of top players, they're very unstable. Here's a graph of expert play rates after our first patch, in the second week and the fifth week. Four characters are highlighted. These numbers are all over the place! I said that expert players pick their characters to win, so why are they moving around so much? Maybe players are figuring out new ways to use these characters post-patch, and they're getting more popular as they win more. But lo! Here are the win rates for the same time period. This graph is pretty zoomed in, too. Almost no variation. So what's going on? Why are play rates changing so much? Well one problem is that it takes players a couple weeks to all gather around an overpowered option, even after its discovered. But the bigger problem, as usual, is free will. In a balanced game there are a lot of ways players can choose characters to win! What's more, these choices exert pressure on each other, so as Kratos gets more popular, Parappa - who has a disadvantage against him - is forced to play less. This creates a potentially never-ending churn of trends in the metagame, even without changes to the game or tactics.



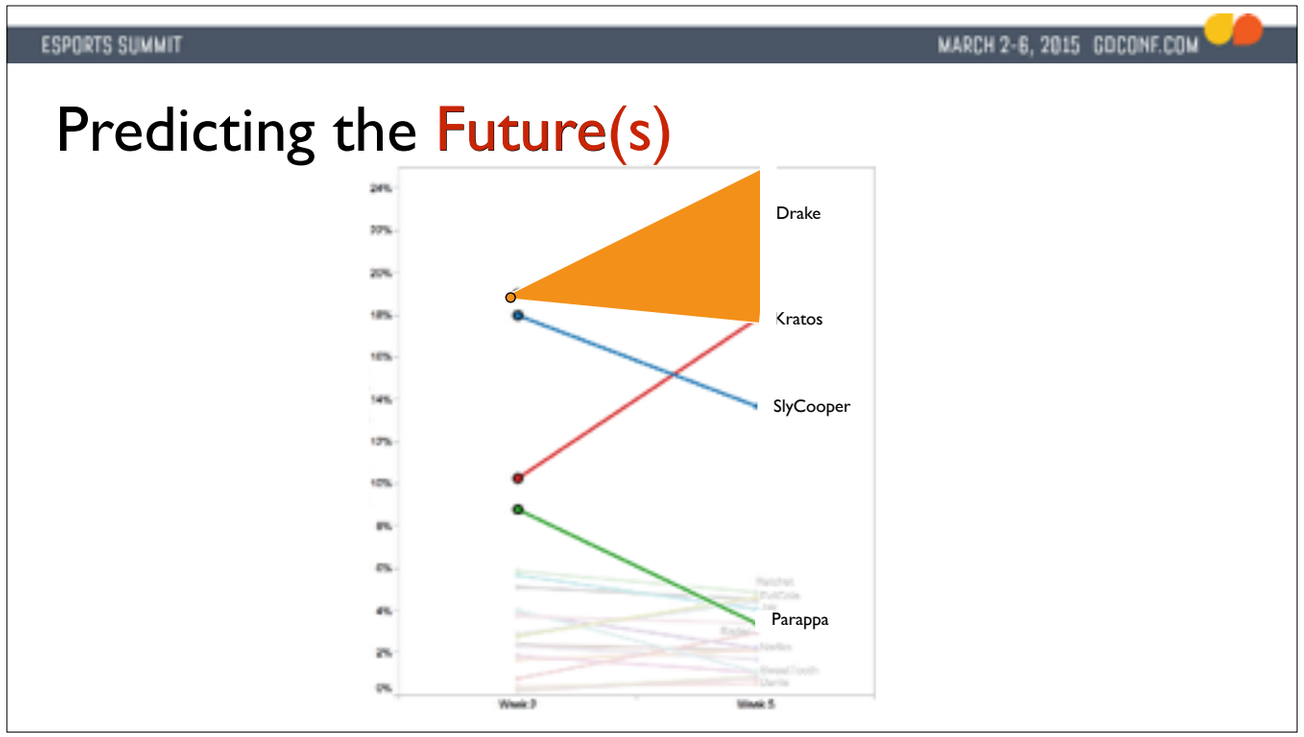
Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!



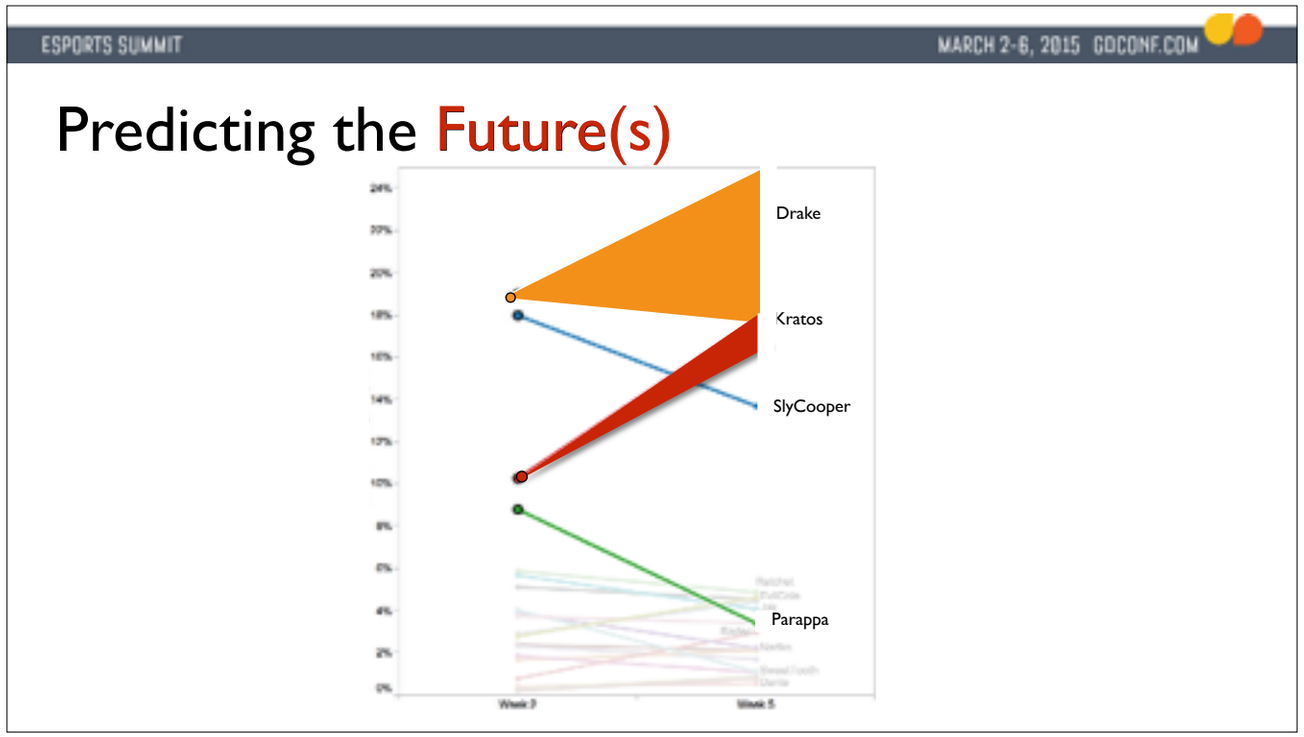
Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!



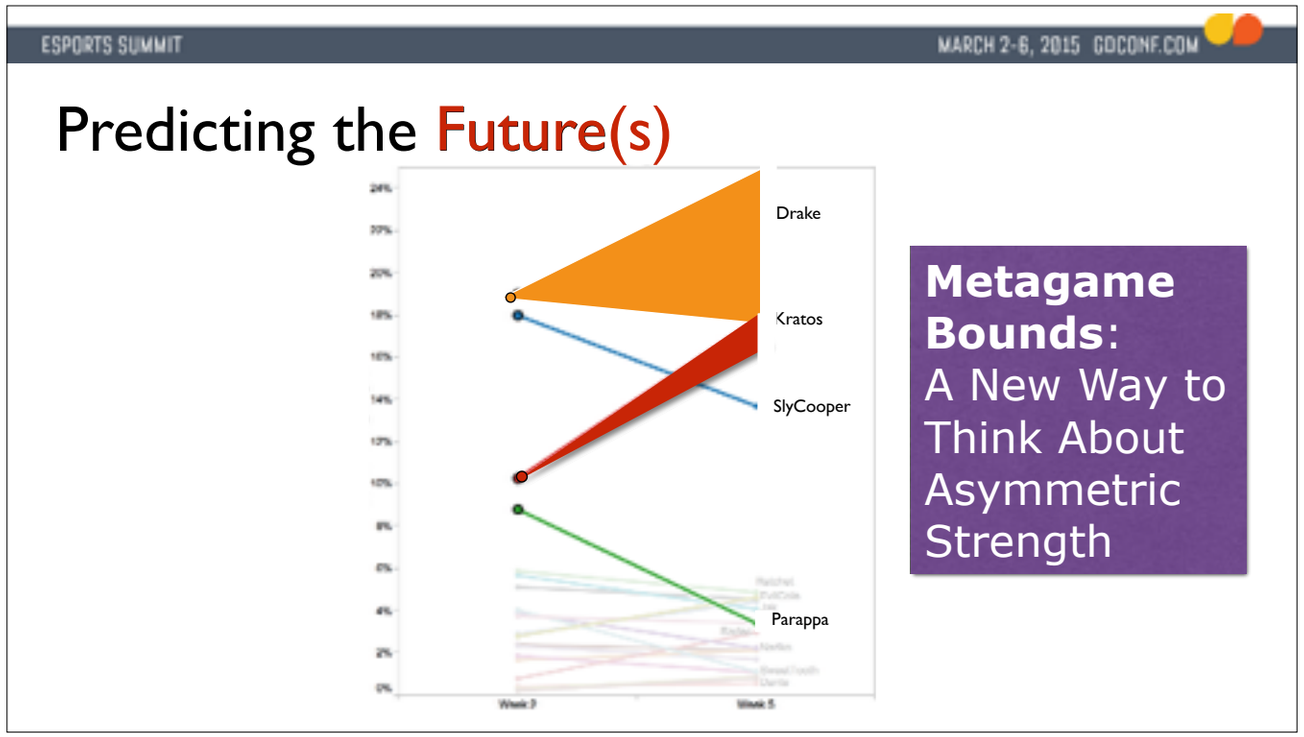
Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!



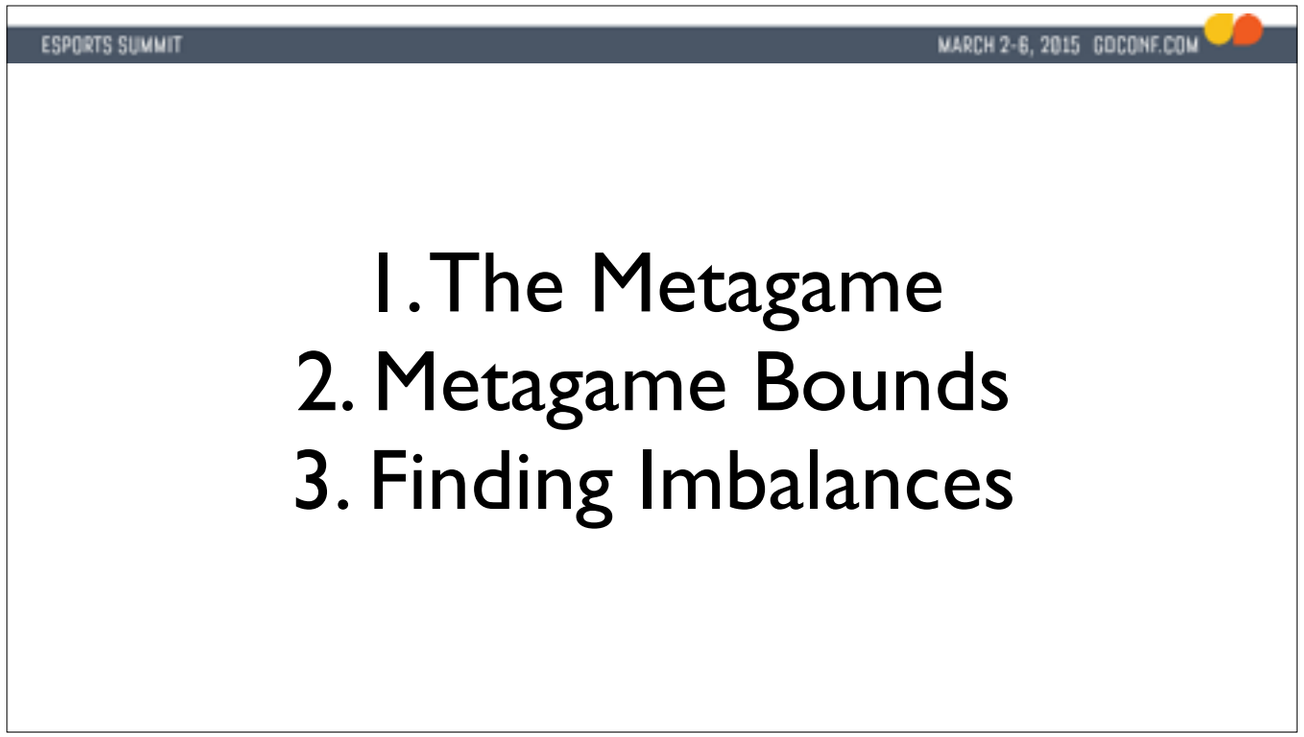
Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!



Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!

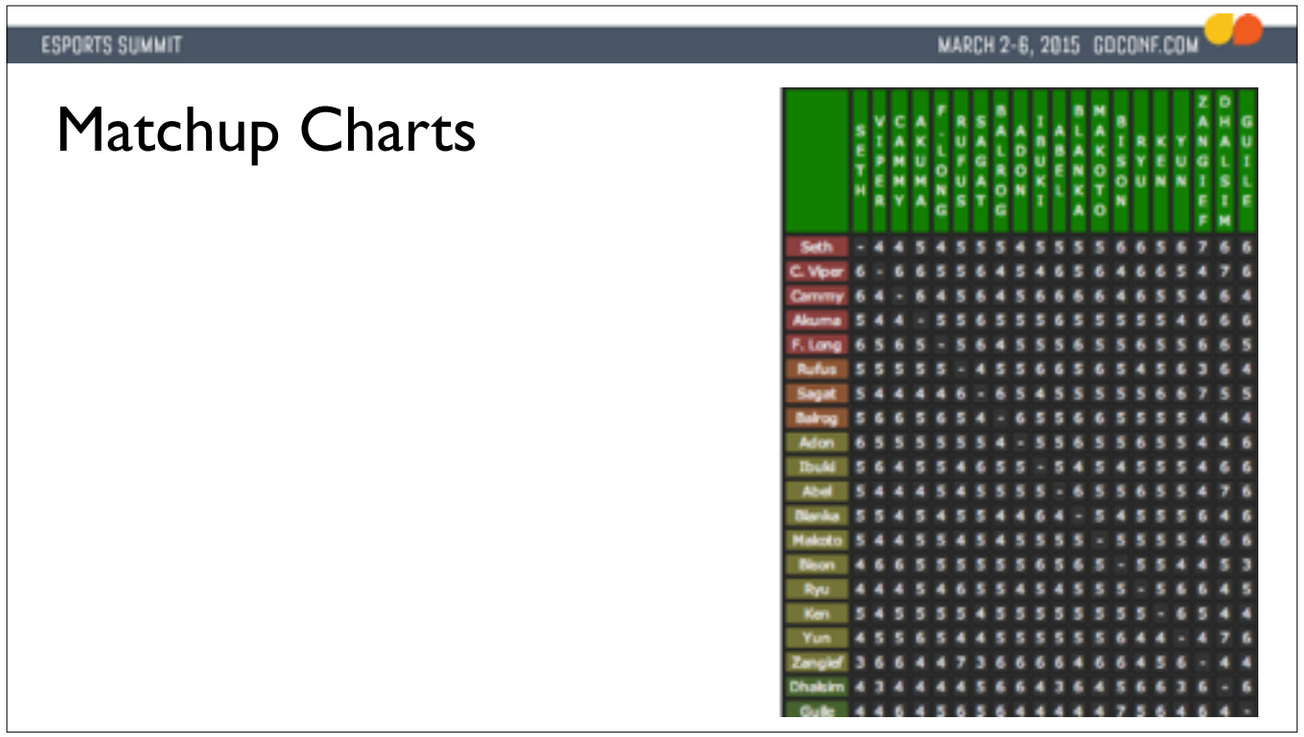


Of course, we'd love it if we could predict the future, and actually know when characters were going to rise to uncomfortable levels of dominance, and potentially adjust the game to stop it before it happens. Like Skynet, without the time travel. At this point it seems pretty unrealistic, right? I've just told you that these play rates could shift all over the place. Drake could rise in popularity like we saw, or stay the same, or drop a bit, or maybe he's more like Sly and drops a ton, or maybe he could even go higher? How are we supposed to predict the future? Well how about this. Let's relax the goal. Let's just say where we *might* end up. What if I could tell you, with some confidence the range of possible frequencies that Drake might reasonably be played by expert players? Or for that matter, that Kratos is essentially bound to eventually be played at a very high frequency? Well, I think that would be a pretty good tool in your tool belt. You could know whether a character might dominate, or disappear from view, or never have a chance of showing up at all, or was even bound to dominate. You could patch after collecting just a couple days of data, rather than waiting for players to explore the space and break the game for you. Hell, if you have enough playtesters, you might not even have to patch at all. This is the vision I'm here to present, and it it is mostly true!



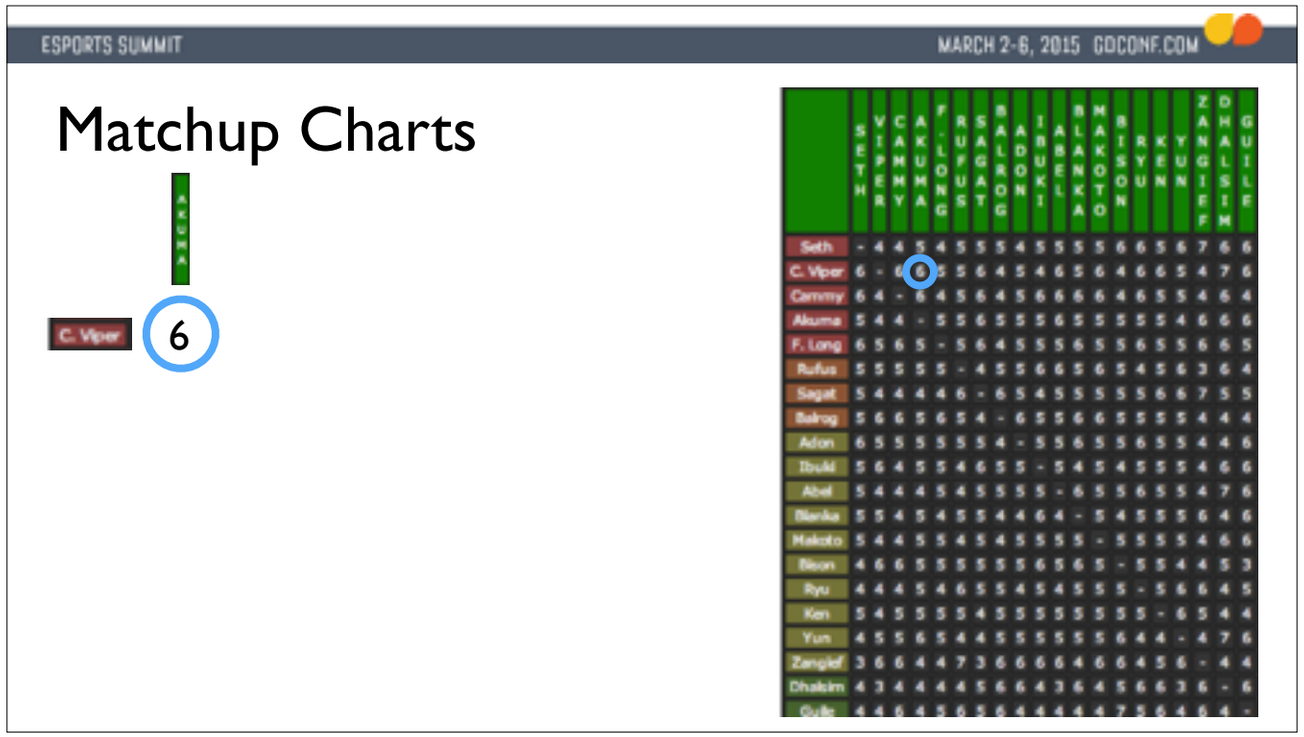
So here are my three sections - they'll come fast! First I'll introduce my slightly weird view of the metagame, and show why it lets us analyze a game in new ways. Then I'll show how that formalization lets us learn the predict the kinds of frequency bounds I just described. Finally I'll give some examples of using metagame bounds more deeply understand the balance of our games.

Okay, onto my view of the metagame.



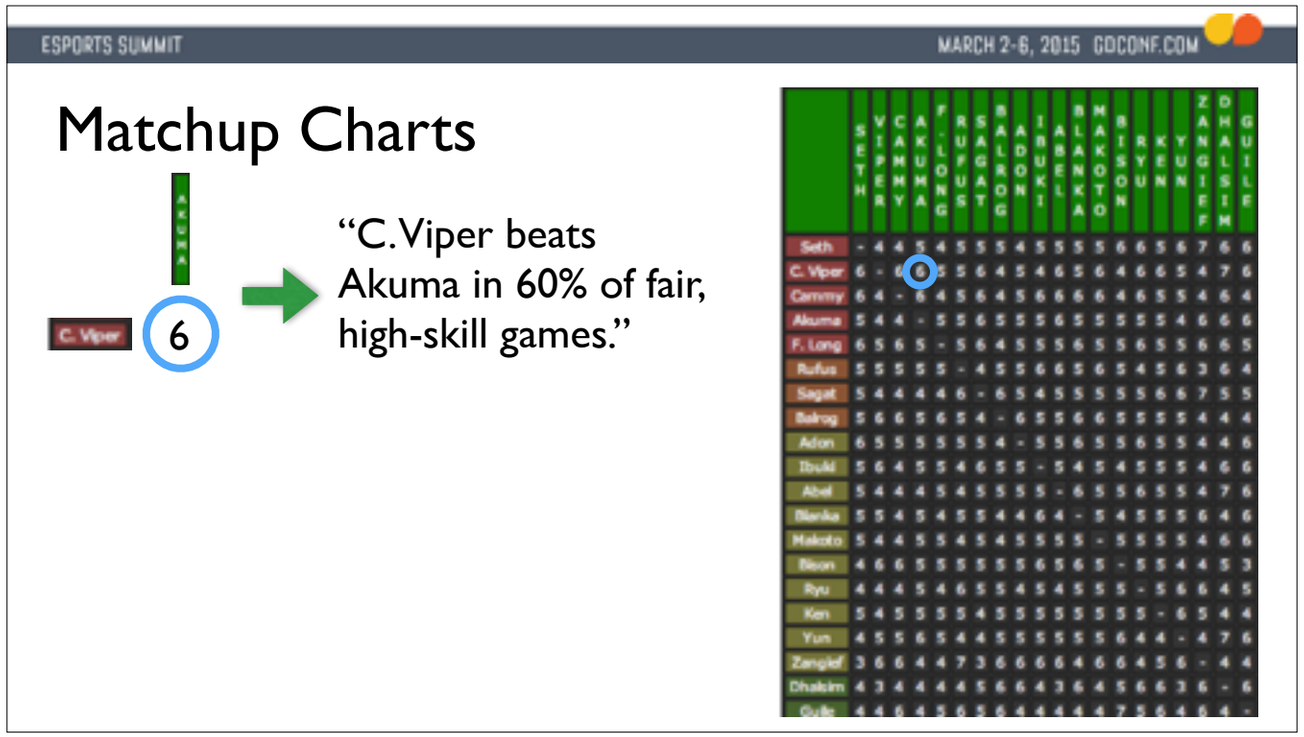
	SETH	VIPER	CAMMY	AKUMA	F.-LONG	RUFUS	SAGAT	BALROG	ADON	IBUKI	ABEL	BLANKA	MAKOTO	BISON	RYU	KEN	YUN	ZANGIEF	DHALSIM	GUILE
Seth	-	4	4	5	4	5	5	5	4	5	5	5	5	6	6	5	6	7	6	6
C. Viper	6	-	6	6	5	5	6	4	5	4	6	5	6	4	6	6	5	4	7	6
Cammy	6	4	-	6	4	5	6	4	5	6	6	6	6	4	6	5	5	4	6	4
Akuma	5	4	4	-	5	5	6	5	5	5	6	5	5	5	5	5	4	6	6	6
F. Long	6	5	6	5	-	5	6	4	5	5	5	6	5	5	6	5	5	6	6	5
Rufus	5	5	5	5	5	-	4	5	5	6	6	5	6	5	4	5	6	3	6	4
Sagat	5	4	4	4	4	6	-	6	5	4	5	5	5	5	5	6	6	7	5	5
Balrog	5	6	6	5	6	5	4	-	6	5	5	6	6	5	5	5	5	4	4	4
Adon	6	5	5	5	5	5	5	4	-	5	5	6	5	5	6	5	5	4	4	6
Ibuki	5	6	4	5	5	4	6	5	5	-	5	4	5	4	5	5	5	4	6	6
Abel	5	4	4	4	5	4	5	5	5	5	-	6	5	5	6	5	5	4	7	6
Blanka	5	5	4	5	4	5	5	4	4	6	4	-	5	4	5	5	5	6	4	6
Makoto	5	4	4	5	5	4	5	4	5	5	5	5	-	5	5	5	5	4	6	6
Bison	4	6	6	5	5	5	5	5	5	6	5	6	5	-	5	5	4	4	5	3
Ryu	4	4	4	5	4	6	5	5	4	5	4	5	5	5	-	5	6	6	4	5
Ken	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	-	6	5	4	4
Yun	4	5	5	6	5	4	4	5	5	5	5	5	5	6	4	4	-	4	7	6
Zangief	3	6	6	4	4	7	3	6	6	6	6	4	6	6	4	5	6	-	4	4
Dhalsim	4	3	4	4	4	4	5	6	6	4	3	6	4	5	6	6	3	6	-	6
Guile	4	4	6	4	5	6	5	6	4	4	4	4	4	7	5	6	4	6	4	-

Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.

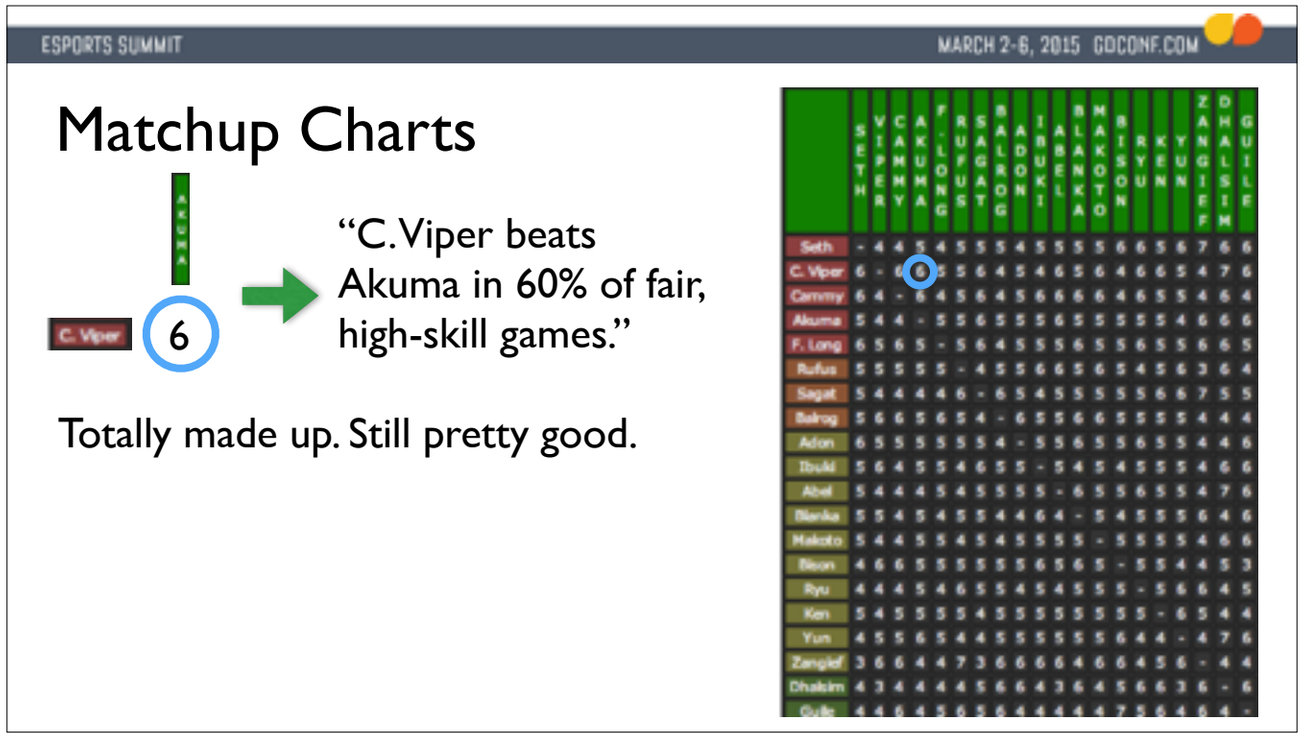


	SETH	VIPER	CAMMY	AKUMA	F.-LONG	RUFUS	SAGAT	BALROG	ADON	IBUKI	ABEL	BLANKA	MAKOTO	BISON	RYU	KEN	YUN	ZANGIEF	DHALSIM	GUILE
Seth	-	4	4	5	4	5	5	5	4	5	5	5	5	6	6	5	6	7	6	6
C. Viper	6	-	6	6	5	5	6	4	5	4	6	5	6	4	6	6	5	4	7	6
Cammy	6	4	-	6	4	5	6	4	5	6	6	6	6	4	6	5	5	4	6	4
Akuma	5	4	4	-	5	5	6	5	5	5	6	5	5	5	5	5	4	6	6	6
F. Long	6	5	6	5	-	5	6	4	5	5	5	6	5	5	6	5	5	6	6	5
Rufus	5	5	5	5	5	-	4	5	5	6	6	5	6	5	4	5	6	3	6	4
Sagat	5	4	4	4	4	6	-	6	5	4	5	5	5	5	5	6	6	7	5	5
Balrog	5	6	6	5	6	5	4	-	6	5	5	6	6	5	5	5	5	4	4	4
Adon	6	5	5	5	5	5	5	4	-	5	5	6	5	5	6	5	5	4	4	6
Ibuki	5	6	4	5	5	4	6	5	5	-	5	4	5	4	5	5	5	4	6	6
Abel	5	4	4	4	5	4	5	5	5	5	-	6	5	5	6	5	5	4	7	6
Blanka	5	5	4	5	4	5	5	4	4	6	4	-	5	4	5	5	5	6	4	6
Makoto	5	4	4	5	5	4	5	4	5	5	5	5	-	5	5	5	5	4	6	6
Bison	4	6	6	5	5	5	5	5	5	6	5	6	5	-	5	5	4	4	5	3
Ryu	4	4	4	5	4	6	5	5	4	5	4	5	5	5	-	5	6	6	4	5
Ken	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	-	6	5	4	4
Yun	4	5	5	6	5	4	4	5	5	5	5	5	5	6	4	4	-	4	7	6
Zangief	3	6	6	4	4	7	3	6	6	6	6	4	6	6	4	5	6	-	4	4
Dhalsim	4	3	4	4	4	4	5	6	6	4	3	6	4	5	6	6	3	6	-	6
Guile	4	4	6	4	5	6	5	6	4	4	4	4	4	7	5	6	4	6	4	-

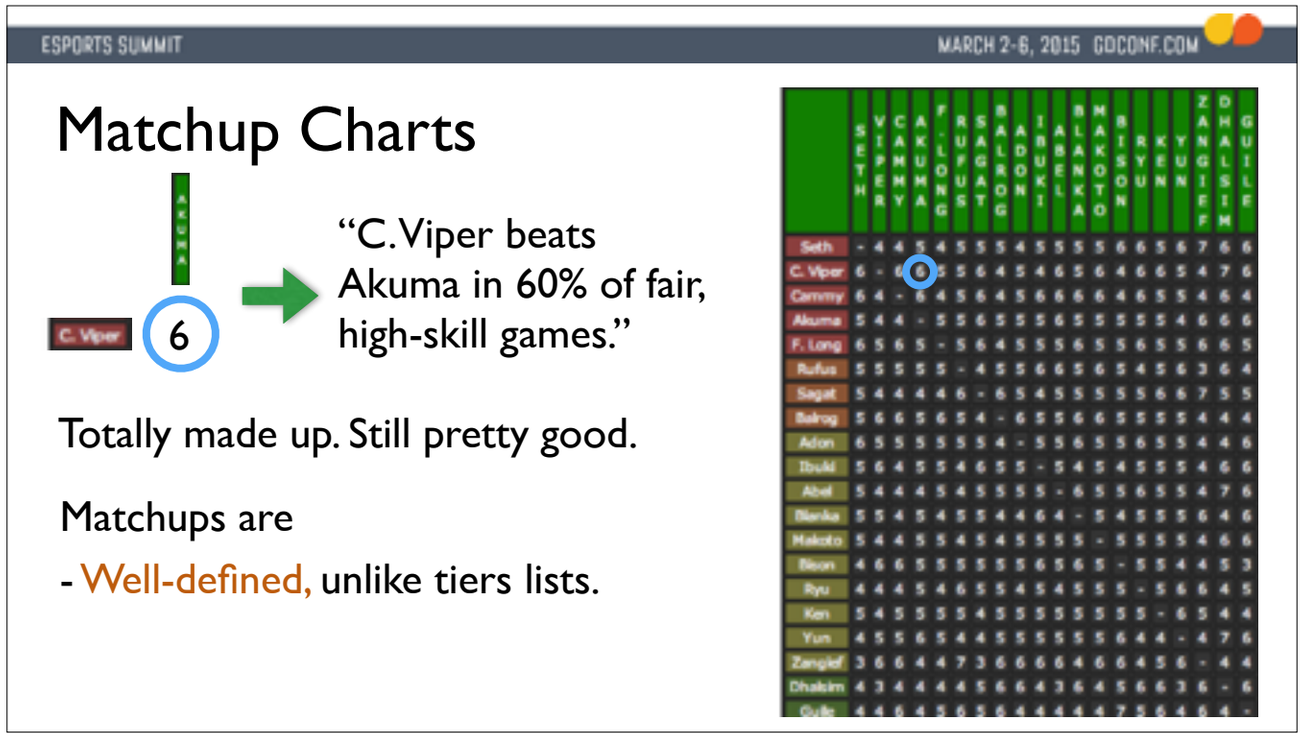
Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.



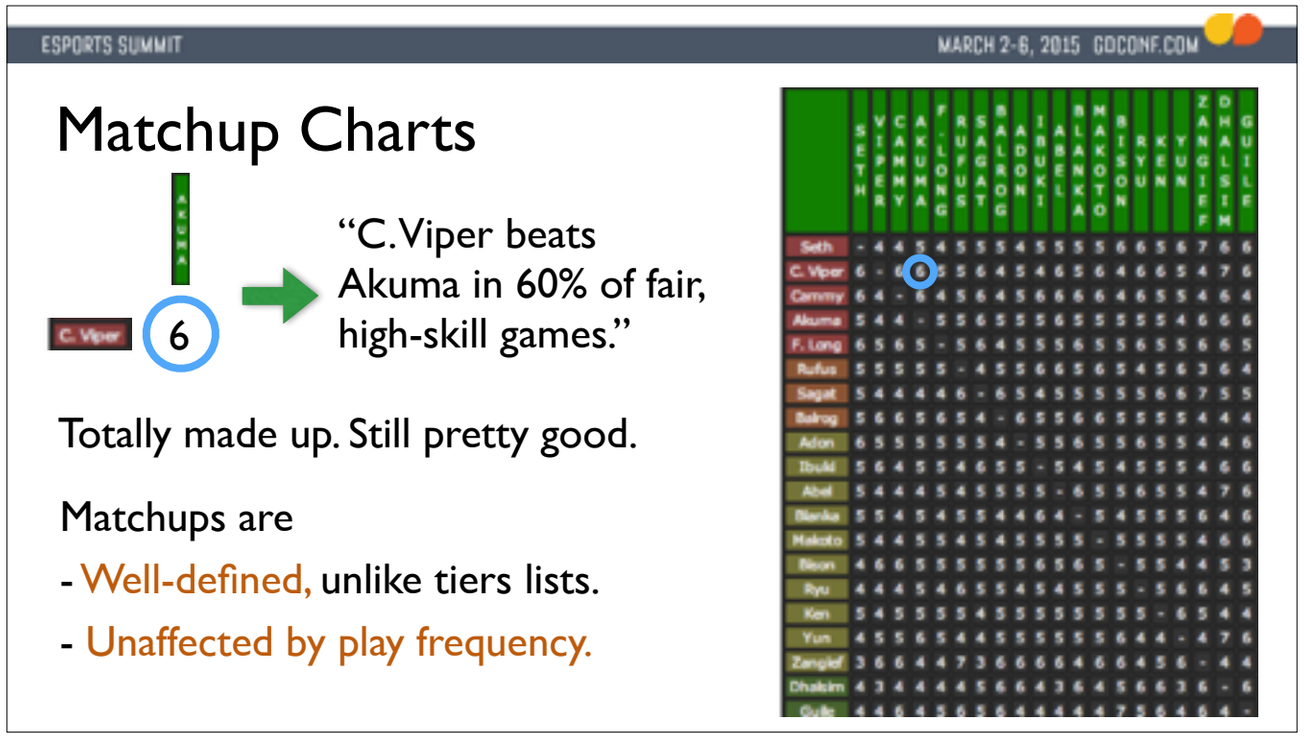
Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.



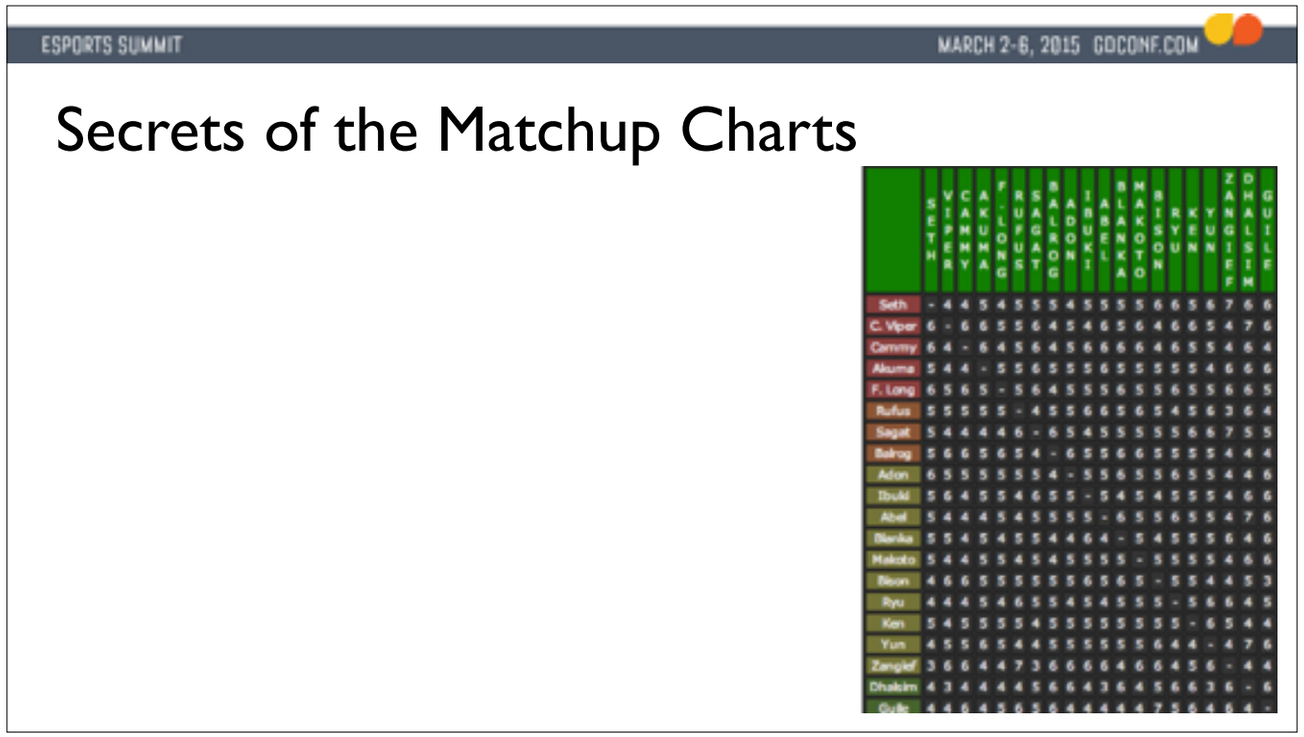
Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.



Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.

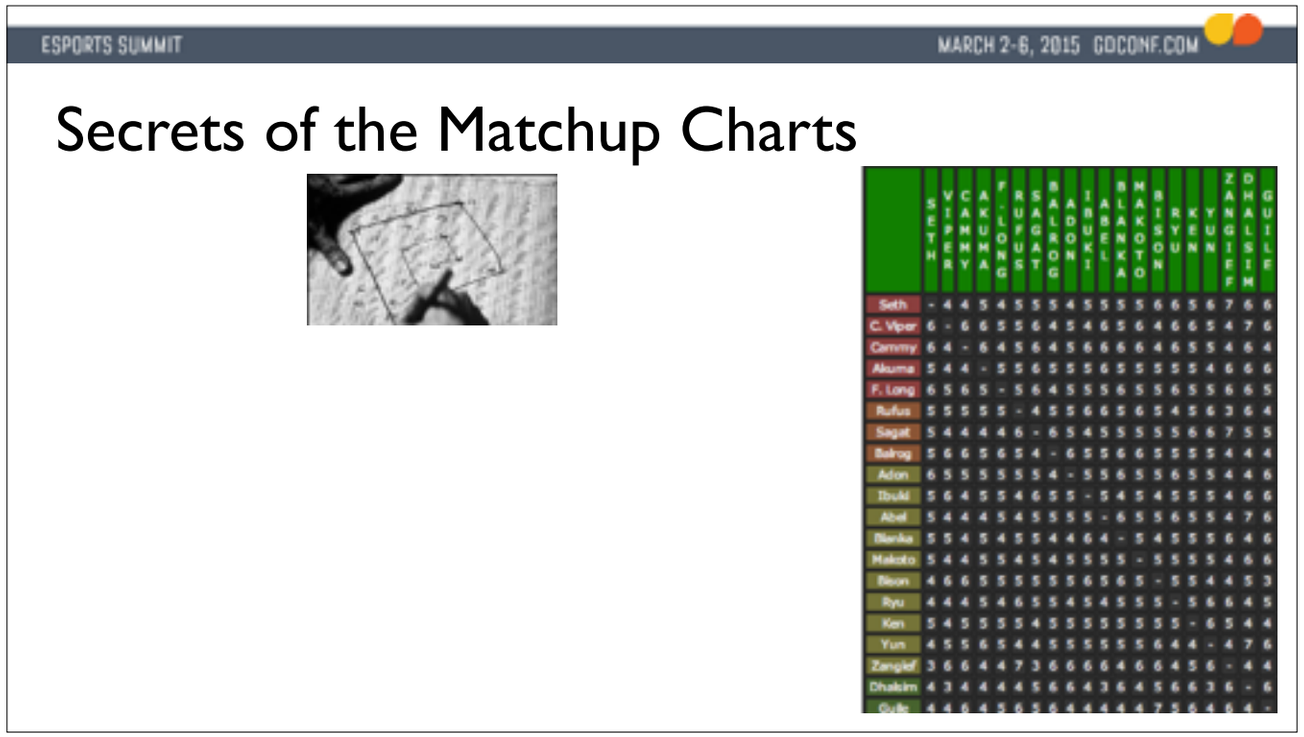


Okay. After working on All-Stars, I went back to the University of Washington, to finish my PhD dissertation, which was, coincidentally, on game balancing. These metagame balance questions kept pestering me. Then one day, I was looking at a matchup chart for Street Fighter 4, depicting the chance that any character should beat any other, in a fair game between skilled players. I.e. if we look at the C. Viper row, and the Akuma column, and we see a 6, that tells us C. Viper beats Akuma in 60% of fair, high-skill games. Now, this particular chart is *totally* made up. It's the collected opinions of some high-level players. That said, I still think it's quite valuable! These numbers are well-defined, unlike tier lists. That alone means that players will be more effective at estimating them. (And what's more we can measure them with data!) They're also mostly unaffected by character popularity, which is the key feature we want. You may object that matchups change over time, as players develop new tech for utilizing characters. True. But I claim that new tech and matchups change slowly compared to the distribution of characters in the metagame, so it's still useful to understand our game for the current state of the matchups.



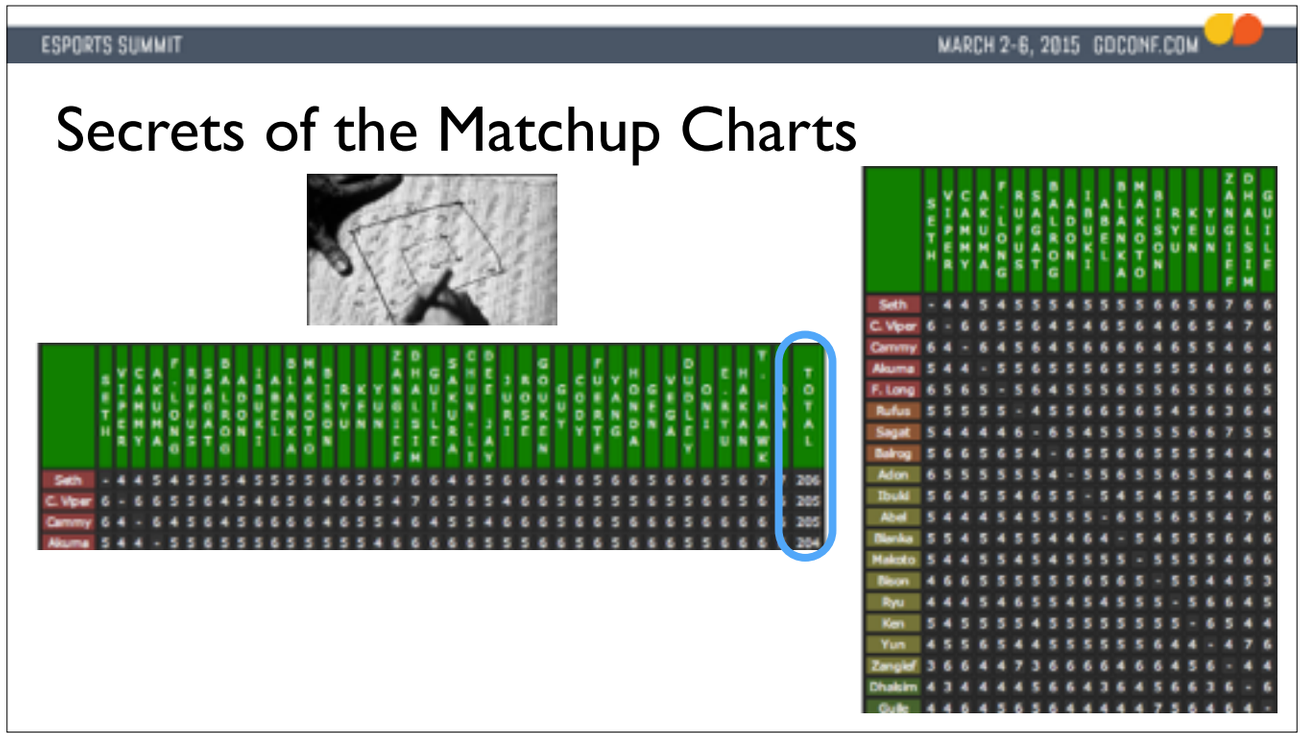
	SETH	VIPER	CAMMY	AKUMA	F.-LONG	RUFUS	SAGAT	BALROG	ADON	IBUKI	ABEL	BLANKA	MAKOTO	BISON	RYU	KEN	YUN	ZANGIEF	DHALSIM	GUILE
Seth	-	4	4	5	4	5	5	5	4	5	5	5	5	6	6	5	6	7	6	6
C. Viper	6	-	6	6	5	5	6	4	5	4	6	5	6	4	6	6	5	4	7	6
Cammy	6	4	-	6	4	5	6	4	5	6	6	6	6	4	6	5	5	4	6	4
Akuma	5	4	4	-	5	5	6	5	5	5	6	5	5	5	5	5	4	6	6	6
F. Long	6	5	6	5	-	5	6	4	5	5	5	6	5	5	6	5	5	6	6	5
Rufus	5	5	5	5	5	-	4	5	5	6	6	5	6	5	4	5	6	3	6	4
Sagat	5	4	4	4	4	6	-	6	5	4	5	5	5	5	5	6	6	7	5	5
Balrog	5	6	6	5	6	5	4	-	6	5	5	6	6	5	5	5	5	4	4	4
Adon	6	5	5	5	5	5	5	4	-	5	5	6	5	5	6	5	5	4	4	6
Ibuki	5	6	4	5	5	4	6	5	5	-	5	4	5	4	5	5	5	4	6	6
Abel	5	4	4	4	5	4	5	5	5	5	-	6	5	5	6	5	5	4	7	6
Blanka	5	5	4	5	4	5	5	4	4	6	4	-	5	4	5	5	5	6	4	6
Makoto	5	4	4	5	5	4	5	4	5	5	5	5	-	5	5	5	5	4	6	6
Bison	4	6	6	5	5	5	5	5	5	6	5	6	5	-	5	5	4	4	5	3
Ryu	4	4	4	5	4	6	5	5	4	5	4	5	5	5	-	5	6	6	4	5
Ken	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	-	6	5	4	4
Yun	4	5	5	6	5	4	4	5	5	5	5	5	5	6	4	4	-	4	7	6
Zangief	3	6	6	4	4	7	3	6	6	6	6	4	6	6	4	5	6	-	4	4
Dhalsim	4	3	4	4	4	4	5	6	6	4	3	6	4	5	6	6	3	6	-	6
Guile	4	4	6	4	5	6	5	6	4	4	4	4	4	7	5	6	4	6	4	-

So my hope was that, like the protagonist in Pi, searching for the secrets of the universe in strings of numbers, we might be able to derive some fundamental truth about a game from this impenetrable mess of numbers. Now I'm not the only one to do this. If you look up any player-created matchup chart online, you'll see this: a sum of all the win rates, for each character. This sum is meant to convey some overall notion character strength. They're ranked by this sum, in fact. And yet, taking sum seemed too simplistic to me. It's equivalent to asking a character's win rate against a uniformly random opponent. Which is even more arbitrary than their win rate against the *current* distribution of characters! So I took this as a challenge. Can we do better? Yes! And we do so by doing something that seems kind of trite: we look at this process of selecting characters, and we take it seriously as a game itself, worthy of its own analysis.

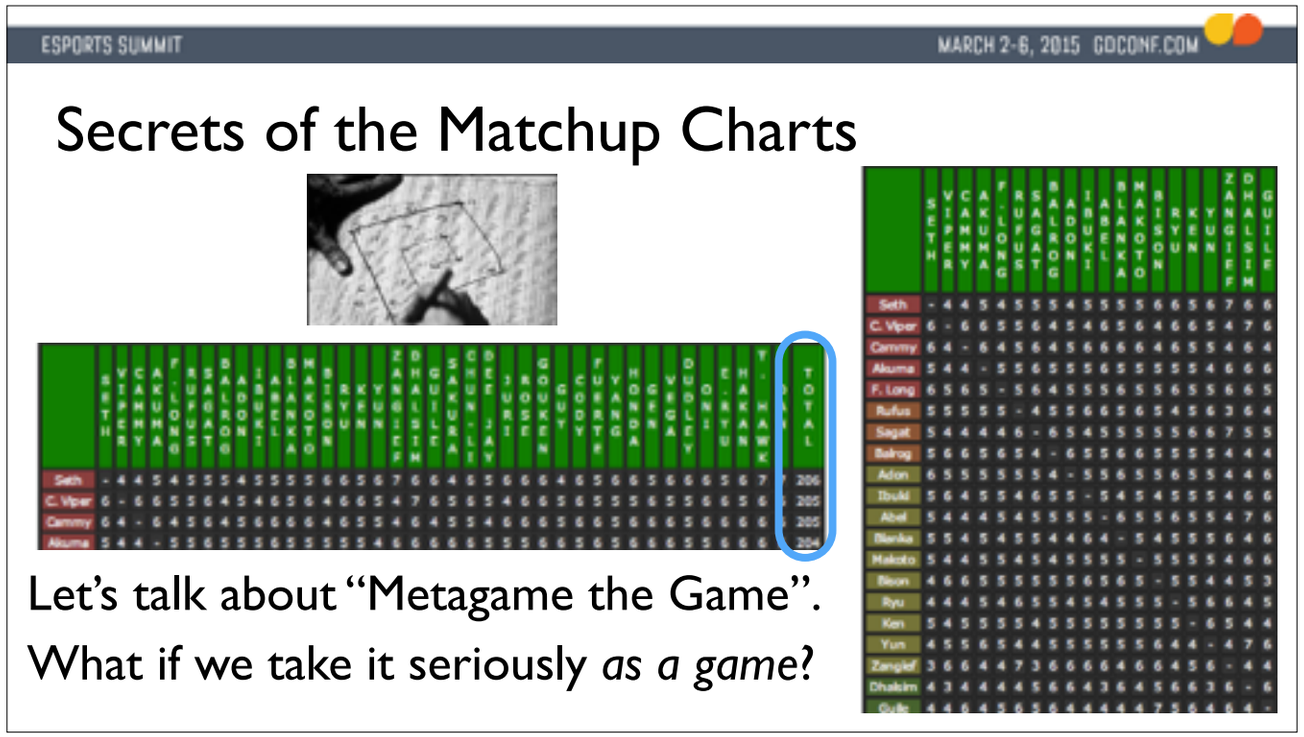


	SETH	VIPER	CAMMY	AKUMA	F.-LONG	RUFUS	SAGAT	BALROG	ADON	IBUKI	ABEL	BLANKA	MAKOTO	BISON	RYU	KEN	YUN	ZANGIEF	DHALSIM	GUILE
Seth	-	4	4	5	4	5	5	5	4	5	5	5	5	6	6	5	6	7	6	6
C. Viper	6	-	6	6	5	5	6	4	5	4	6	5	6	4	6	6	5	4	7	6
Cammy	6	4	-	6	4	5	6	4	5	6	6	6	6	4	6	5	5	4	6	4
Akuma	5	4	4	-	5	5	6	5	5	5	6	5	5	5	5	5	4	6	6	6
F. Long	6	5	6	5	-	5	6	4	5	5	5	6	5	5	6	5	5	6	6	5
Rufus	5	5	5	5	5	-	4	5	5	6	6	5	6	5	4	5	6	3	6	4
Sagat	5	4	4	4	4	6	-	6	5	4	5	5	5	5	5	6	6	7	5	5
Balrog	5	6	6	5	6	5	4	-	6	5	5	6	6	5	5	5	5	4	4	4
Adon	6	5	5	5	5	5	5	4	-	5	5	6	5	5	6	5	5	4	4	6
Ibuki	5	6	4	5	5	4	6	5	5	-	5	4	5	4	5	5	5	4	6	6
Abel	5	4	4	4	5	4	5	5	5	5	-	6	5	5	6	5	5	4	7	6
Blanka	5	5	4	5	4	5	5	4	4	6	4	-	5	4	5	5	5	6	4	6
Makoto	5	4	4	5	5	4	5	4	5	5	5	5	-	5	5	5	5	4	6	6
Bison	4	6	6	5	5	5	5	5	5	6	5	6	5	-	5	5	4	4	5	3
Ryu	4	4	4	5	4	6	5	5	4	5	4	5	5	5	-	5	6	6	4	5
Ken	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	-	6	5	4	4
Yun	4	5	5	6	5	4	4	5	5	5	5	5	5	6	4	4	-	4	7	6
Zangief	3	6	6	4	4	7	3	6	6	6	6	4	6	6	4	5	6	-	4	4
Dhalsim	4	3	4	4	4	4	5	6	6	4	3	6	4	5	6	6	3	6	-	6
Guile	4	4	6	4	5	6	5	6	4	4	4	4	4	7	5	6	4	6	4	-

So my hope was that, like the protagonist in Pi, searching for the secrets of the universe in strings of numbers, we might be able to derive some fundamental truth about a game from this impenetrable mess of numbers. Now I'm not the only one to do this. If you look up any player-created matchup chart online, you'll see this: a sum of all the win rates, for each character. This sum is meant to convey some overall notion character strength. They're ranked by this sum, in fact. And yet, taking sum seemed too simplistic to me. It's equivalent to asking a character's win rate against a uniformly random opponent. Which is even more arbitrary than their win rate against the *current* distribution of characters! So I took this as a challenge. Can we do better? Yes! And we do so by doing something that seems kind of trite: we look at this process of selecting characters, and we take it seriously as a game itself, worthy of its own analysis.



So my hope was that, like the protagonist in Pi, searching for the secrets of the universe in strings of numbers, we might be able to derive some fundamental truth about a game from this impenetrable mess of numbers. Now I'm not the only one to do this. If you look up any player-created matchup chart online, you'll see this: a sum of all the win rates, for each character. This sum is meant to convey some overall notion character strength. They're ranked by this sum, in fact. And yet, taking sum seemed too simplistic to me. It's equivalent to asking a character's win rate against a uniformly random opponent. Which is even more arbitrary than their win rate against the *current* distribution of characters! So I took this as a challenge. Can we do better? Yes! And we do so by doing something that seems kind of trite: we look at this process of selecting characters, and we take it seriously as a game itself, worthy of its own analysis.



So my hope was that, like the protagonist in Pi, searching for the secrets of the universe in strings of numbers, we might be able to derive some fundamental truth about a game from this impenetrable mess of numbers. Now I'm not the only one to do this. If you look up any player-created matchup chart online, you'll see this: a sum of all the win rates, for each character. This sum is meant to convey some overall notion character strength. They're ranked by this sum, in fact. And yet, taking sum seemed too simplistic to me. It's equivalent to asking a character's win rate against a uniformly random opponent. Which is even more arbitrary than their win rate against the *current* distribution of characters! So I took this as a challenge. Can we do better? Yes! And we do so by doing something that seems kind of trite: we look at this process of selecting characters, and we take it seriously as a game itself, worthy of its own analysis.

Call it "The Selection Game".

You Pick:

I Pick:

	Ryu	Zangief	Guile
Ryu			
Zangief			
Guile			

Let's call this "metagame game" the "Selection Game", just to be clear. We sit down to play a game. We each close our eyes and choose a character in secret. But we don't play the game. Instead, robots programmed to play exactly like us play in our stead. So all I care about is choosing the best character, after which I win with some probability. (Ties are folded in by treating them as a 50% chance of winning.) This game may sound dumb, but I promise you it is interesting. I also know this doesn't capture the full experience of the metagame. But I'm going to make an argument pretty soon that it does so better than you might, think, so be patient with me.

Call it "The Selection Game".

You Pick:

I Pick:

	Ryu	Zangief	Guile
Ryu	50%	70%	40%
Zangief	30%	50%	60%
Guile	60%	40%	50%

Let's call this "metagame game" the "Selection Game", just to be clear. We sit down to play a game. We each close our eyes and choose a character in secret. But we don't play the game. Instead, robots programmed to play exactly like us play in our stead. So all I care about is choosing the best character, after which I win with some probability. (Ties are folded in by treating them as a 50% chance of winning.) This game may sound dumb, but I promise you it is interesting. I also know this doesn't capture the full experience of the metagame. But I'm going to make an argument pretty soon that it does so better than you might, think, so be patient with me.

Call it "The Selection Game".

You Pick:

I Pick:

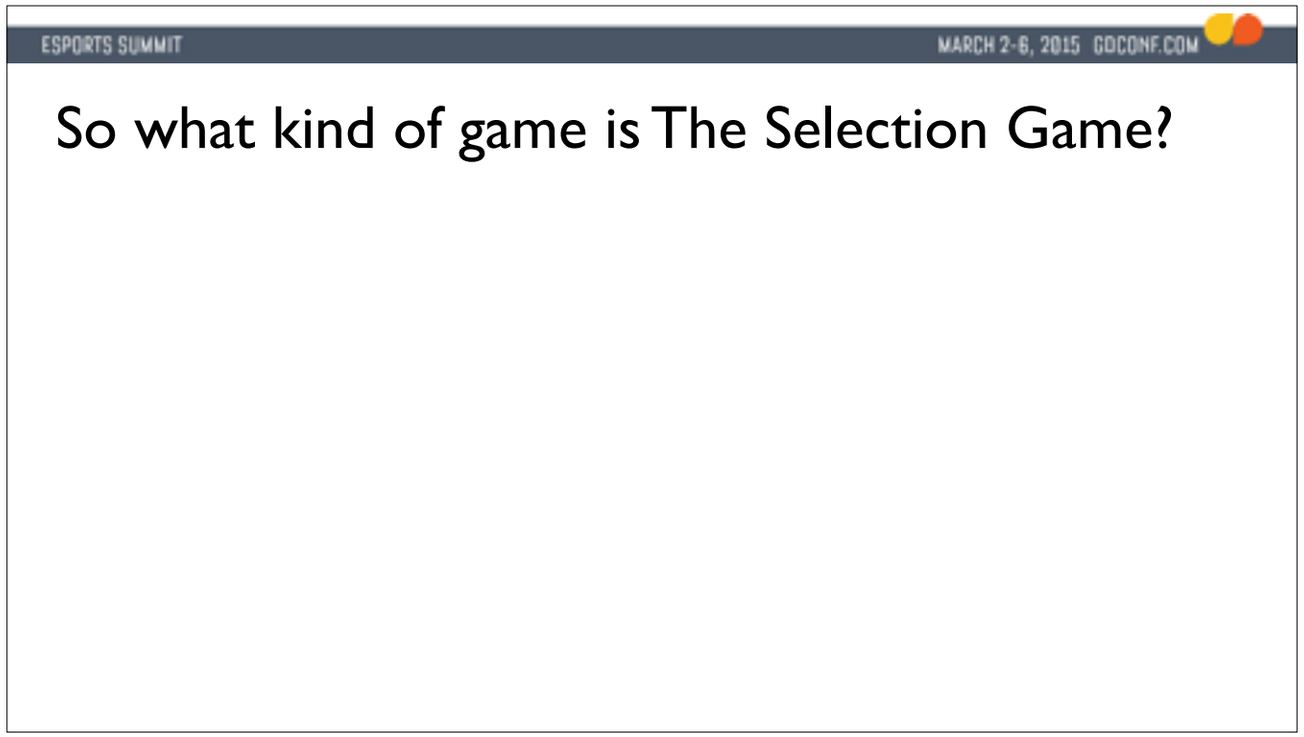
	Ryu	Zangief	Guile
Ryu	50%	70%	40%
Zangief	30%	50%	60%
Guile	60%	40%	50%

Robots play.

60% means ***I*** win 60% of those games. ***You*** win 40%.

(We treat ties as 50% win.)

Let's call this "metagame game" the "Selection Game", just to be clear. We sit down to play a game. We each close our eyes and choose a character in secret. But we don't play the game. Instead, robots programmed to play exactly like us play in our stead. So all I care about is choosing the best character, after which I win with some probability. (Ties are folded in by treating them as a 50% chance of winning.) This game may sound dumb, but I promise you it is interesting. I also know this doesn't capture the full experience of the metagame. But I'm going to make an argument pretty soon that it does so better than you might, think, so be patient with me.



What kind of a game is this? Well, it's basically Rock-Paper-Scissors, though with more interesting variation. In RPS we each choose a move simultaneously, and one of us wins, depending on our choices. It's a game about Yomi - about understanding your opponent's mindset. Fortunately, rock-paper-scissors is a game we understand *very* thoroughly cousins. That's what game theory is about, and I need to explain exactly one game theory concept to you. An optimal strategy is a strategy that guarantees you can win half your games. Of course, no simple strategy in RPS accomplishes this: rock can be beaten every time by paper, and so on. To get an optimal strategy in RPS, you need randomness.

So what kind of game is The Selection Game?

It's kind of Rock-Paper-Scissors.

	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

What kind of a game is this? Well, it's basically Rock-Paper-Scissors, though with more interesting variation. In RPS we each choose a move simultaneously, and one of us wins, depending on our choices. It's a game about Yomi - about understanding your opponent's mindset. Fortunately, rock-paper-scissors is a game we understand *very* thoroughly cousins. That's what game theory is about, and I need to explain exactly one game theory concept to you. An optimal strategy is a strategy that guarantees you can win half your games. Of course, no simple strategy in RPS accomplishes this: rock can be beaten every time by paper, and so on. To get an optimal strategy in RPS, you need randomness.

So what kind of game is The Selection Game?

It's kind of Rock-Paper-Scissors.

Game Theory

	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

What kind of a game is this? Well, it's basically Rock-Paper-Scissors, though with more interesting variation. In RPS we each choose a move simultaneously, and one of us wins, depending on our choices. It's a game about Yomi - about understanding your opponent's mindset. Fortunately, rock-paper-scissors is a game we understand *very* thoroughly cousins. That's what game theory is about, and I need to explain exactly one game theory concept to you. An optimal strategy is a strategy that guarantees you can win half your games. Of course, no simple strategy in RPS accomplishes this: rock can be beaten every time by paper, and so on. To get an optimal strategy in RPS, you need randomness.

So what kind of game is The Selection Game?

It's kind of Rock-Paper-Scissors.

Game Theory

	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

Optimal Strategy: wins at least half of games, no matter what opponent knows or does.

What kind of a game is this? Well, it's basically Rock-Paper-Scissors, though with more interesting variation. In RPS we each choose a move simultaneously, and one of us wins, depending on our choices. It's a game about Yomi - about understanding your opponent's mindset. Fortunately, rock-paper-scissors is a game we understand *very* thoroughly cousins. That's what game theory is about, and I need to explain exactly one game theory concept to you. An optimal strategy is a strategy that guarantees you can win half your games. Of course, no simple strategy in RPS accomplishes this: rock can be beaten every time by paper, and so on. To get an optimal strategy in RPS, you need randomness.

So what kind of game is The Selection Game?

It's kind of Rock-Paper-Scissors.

Game Theory

	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

Any fixed strategy can be exploited. Go random.

Optimal Strategy: wins at least half of games, no matter what opponent knows or does.

What kind of a game is this? Well, it's basically Rock-Paper-Scissors, though with more interesting variation. In RPS we each choose a move simultaneously, and one of us wins, depending on our choices. It's a game about Yomi - about understanding your opponent's mindset. Fortunately, rock-paper-scissors is a game we understand *very* thoroughly cousins. That's what game theory is about, and I need to explain exactly one game theory concept to you. An optimal strategy is a strategy that guarantees you can win half your games. Of course, no simple strategy in RPS accomplishes this: rock can be beaten every time by paper, and so on. To get an optimal strategy in RPS, you need randomness.

Mixed Strategies

Me \ You	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

So let's consider a mixed strategy, which is a probability distribution over "pure strategies". Say I play each of rock, paper, and scissors 1/3 of the time. Then you're living in this hazy world, where it's not clear which part of the game you're playing in. If you play Rock, you'll only win half the time. The same is true if you play paper, or scissors, or even any random combination of them. Uniformly random wins half the time even if you know that I'm doing it and try to exploit me. That's optimal because the game is symmetric – I can't possibly guarantee more than half wins. Note that this notion of optimal is a bit restricted: I can play even better if I know your weaknesses. But optimality is more like a guarantee: at least half wins.

Mixed Strategies

A mixed strategy is a probability distribution over "pure strategies".

Me \ You	Rock	Paper	Scissors
Rock	50%	0%	100%
Paper	100%	50%	0%
Scissors	0%	100%	50%

So let's consider a mixed strategy, which is a probability distribution over "pure strategies". Say I play each of rock, paper, and scissors 1/3 of the time. Then you're living in this hazy world, where it's not clear which part of the game you're playing in. If you play Rock, you'll only win half the time. The same is true if you play paper, or scissors, or even any random combination of them. Uniformly random wins half the time even if you know that I'm doing it and try to exploit me. That's optimal because the game is symmetric – I can't possibly guarantee more than half wins. Note that this notion of optimal is a bit restricted: I can play even better if I know your weaknesses. But optimality is more like a guarantee: at least half wins.

Mixed Strategies

A mixed strategy is a probability distribution over "pure strategies".

	Me \ You	Rock	Paper	Scissors
33%	Rock	50%	0%	100%
33%	Paper	100%	50%	0%
33%	Scissors	0%	100%	50%

So let's consider a mixed strategy, which is a probability distribution over "pure strategies". Say I play each of rock, paper, and scissors 1/3 of the time. Then you're living in this hazy world, where it's not clear which part of the game you're playing in. If you play Rock, you'll only win half the time. The same is true if you play paper, or scissors, or even any random combination of them. Uniformly random wins half the time even if you know that I'm doing it and try to exploit me. That's optimal because the game is symmetric – I can't possibly guarantee more than half wins. Note that this notion of optimal is a bit restricted: I can play even better if I know your weaknesses. But optimality is more like a guarantee: at least half wins.

Mixed Strategies

A mixed strategy is a probability distribution over "pure strategies".

	Me \ You	Rock	Paper	Scissors
33%	Rock	50%	0%	100%
33%	Paper	100%	50%	0%
33%	Scissors	0%	100%	50%

So let's consider a mixed strategy, which is a probability distribution over "pure strategies". Say I play each of rock, paper, and scissors 1/3 of the time. Then you're living in this hazy world, where it's not clear which part of the game you're playing in. If you play Rock, you'll only win half the time. The same is true if you play paper, or scissors, or even any random combination of them. Uniformly random wins half the time even if you know that I'm doing it and try to exploit me. That's optimal because the game is symmetric – I can't possibly guarantee more than half wins. Note that this notion of optimal is a bit restricted: I can play even better if I know your weaknesses. But optimality is more like a guarantee: at least half wins.

Mixed Strategies

A mixed strategy is a probability distribution over "pure strategies".

	Me \ You	Rock	Paper	Scissors
33%	Rock	50%	0%	100%
33%	Paper	100%	50%	0%
33%	Scissors	0%	100%	50%

Uniformly random wins at least half the time.
Optimal because game is symmetric!

So let's consider a mixed strategy, which is a probability distribution over "pure strategies". Say I play each of rock, paper, and scissors 1/3 of the time. Then you're living in this hazy world, where it's not clear which part of the game you're playing in. If you play Rock, you'll only win half the time. The same is true if you play paper, or scissors, or even any random combination of them. Uniformly random wins half the time even if you know that I'm doing it and try to exploit me. That's optimal because the game is symmetric – I can't possibly guarantee more than half wins. Note that this notion of optimal is a bit restricted: I can play even better if I know your weaknesses. But optimality is more like a guarantee: at least half wins.

Now that you understand what I mean by the metagame today, we're in a position to infer some understanding of how real people might play.

Play Rates from Nash Equilbria

If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.

Play Rates from Nash Equilbria

	Terran	Protoss	Zerg
Terran	50%	59%	56%
Protoss	41%	50%	47%
Zerg	44%	53%	50%

If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.

Play Rates from Nash Equilbria

		Terran	Protoss	Zerg
?%	Terran	50%	59%	56%
?%	Protoss	41%	50%	47%
?%	Zerg	44%	53%	50%

If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.

Play Rates from Nash Equilbria

		Terran	Protoss	Zerg
?%	Terran	50%	59%	56%
?%	Protoss	41%	50%	47%
?%	Zerg	44%	53%	50%

Linear Program

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{im}x_i\right)\right\}$$

$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

$$x_i \geq 0 \qquad \forall i$$

If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.

Play Rates from Nash Equilbria

		Terran	Protoss	Zerg
36%	Terran	50%	59%	56%
34%	Protoss	41%	50%	47%
30%	Zerg	44%	53%	50%

Linear Program

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{im}x_i\right)\right\}$$

$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

$$x_i \geq 0 \qquad \forall i$$

If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.

Play Rates from Nash Equilbria

Expected play frequencies?
Two big problems.

		Terran	Protoss	Zerg
36%	Terran	50%	59%	56%
34%	Protoss	41%	50%	47%
30%	Zerg	44%	53%	50%

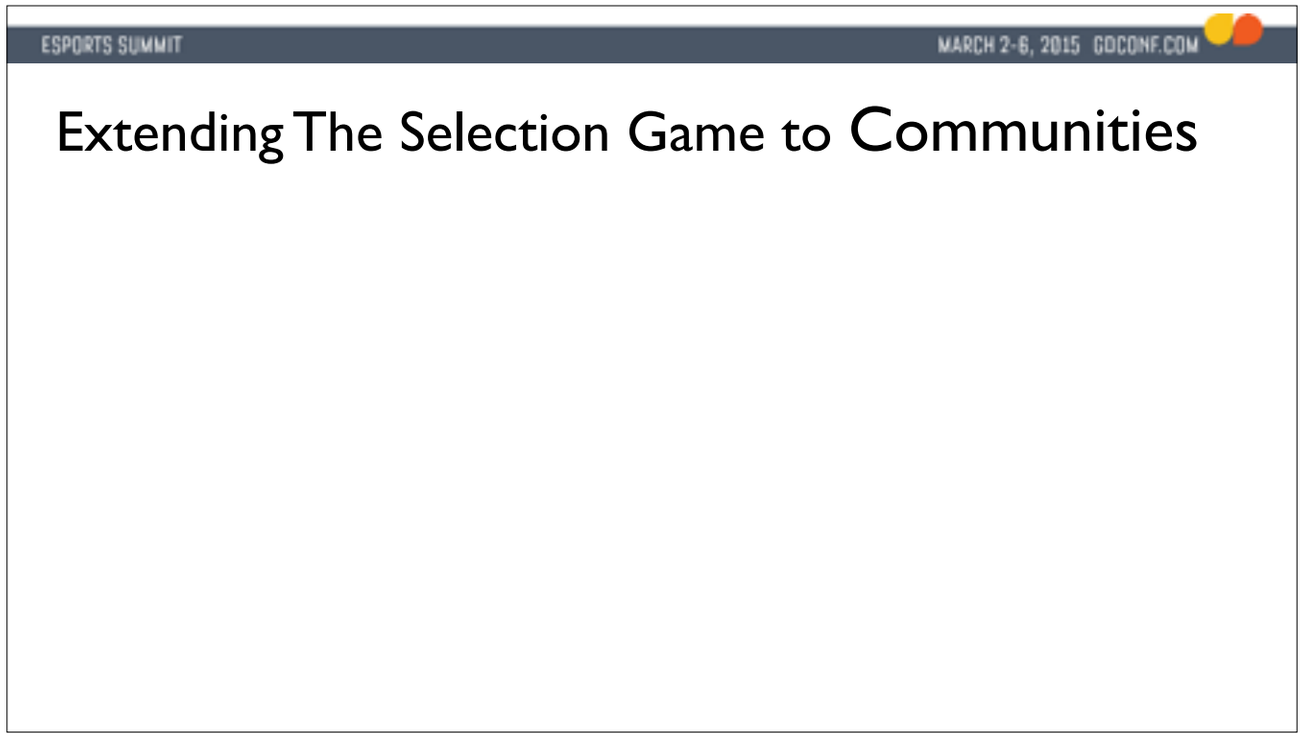
Linear Program

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{im}x_i\right)\right\}$$

$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

$$x_i \geq 0 \qquad \forall i$$

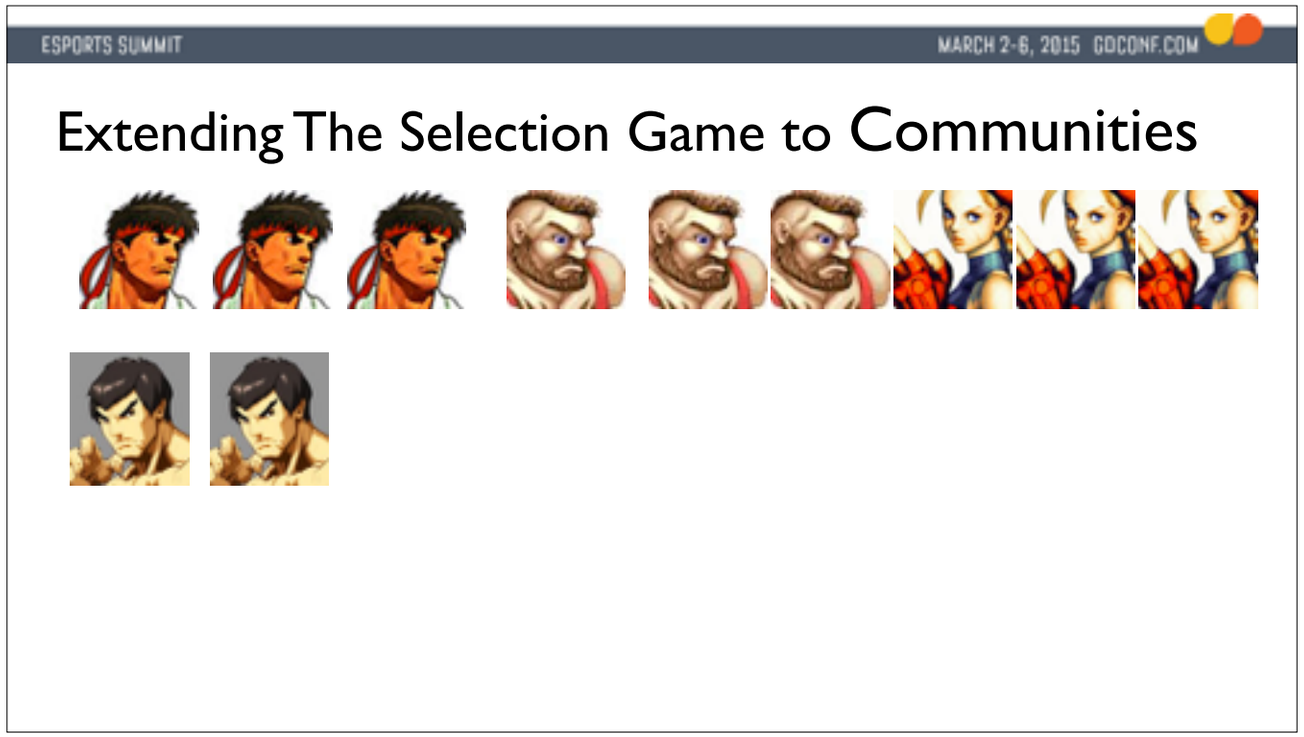
If you've been following closely, you might have an inkling about what I'm going to do. Let me start with a straw man. Consider these win rates, sampled from some Starcraft 2 tournaments a couple years ago. Well we know there exists some optimal strategy for selecting which race to play. But how do we actually find it? I've got some good news for you buddy. There are several well-known algorithms for finding the optimal strategy of a zero-sum game, quite efficiently. You don't have to understand the details, but one simple way is to express the problem as a linear program - basically a set of linear constraints - then apply a general-purpose linear program solver. Run that solver, and very quickly you get something like these probabilities! Voila, we know the expected frequency that skilled players should play each race! Or… not. Two big problems.



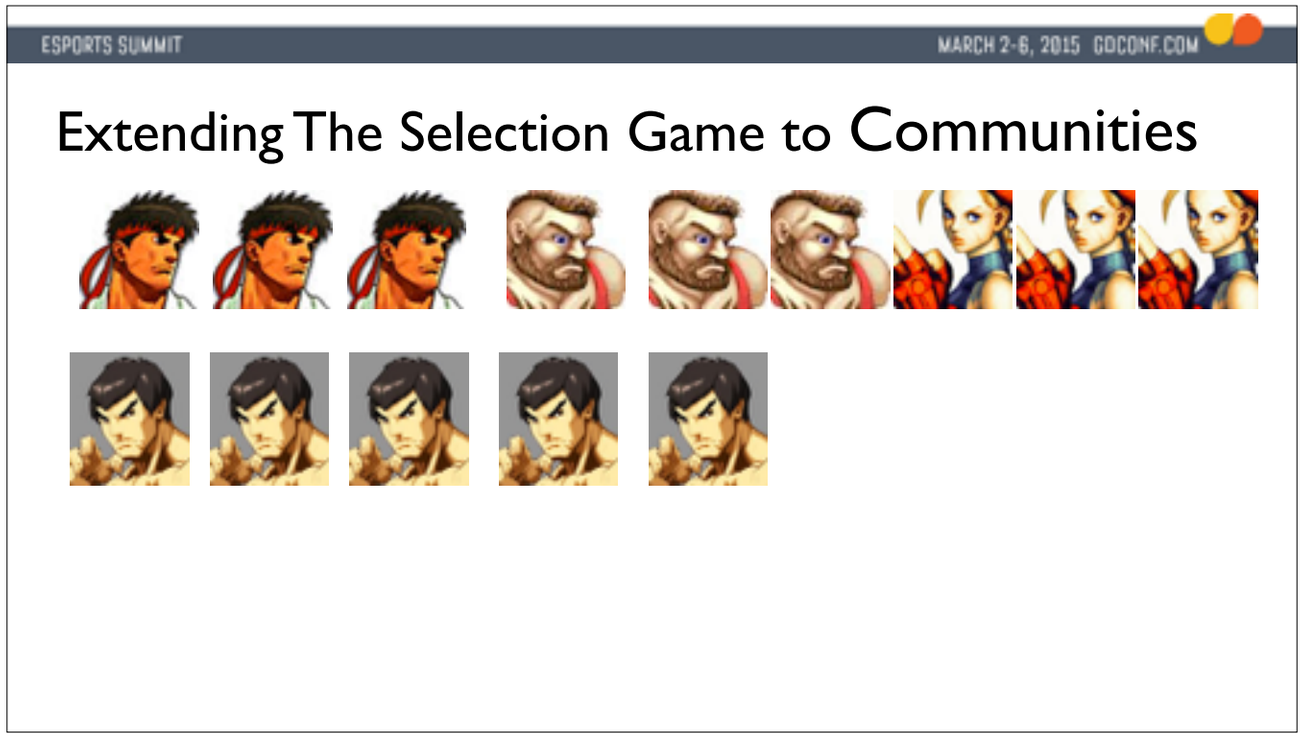
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!

Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!

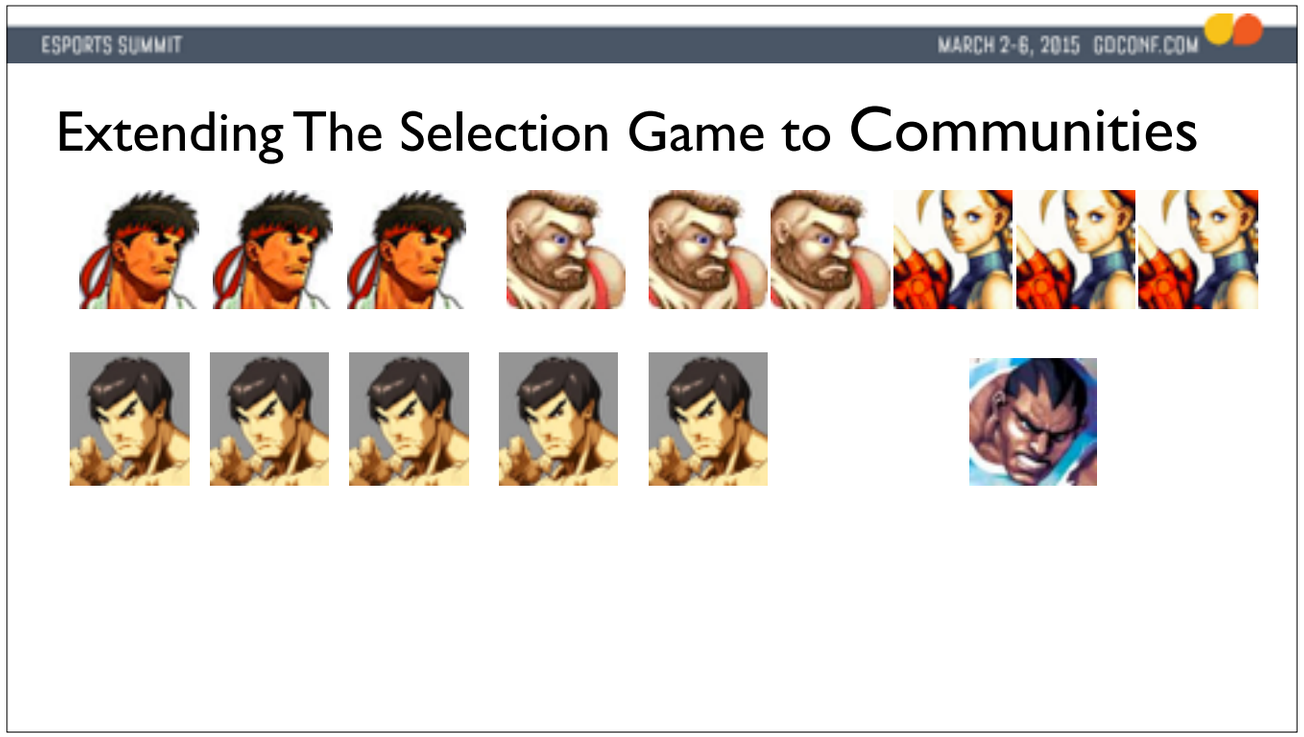
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



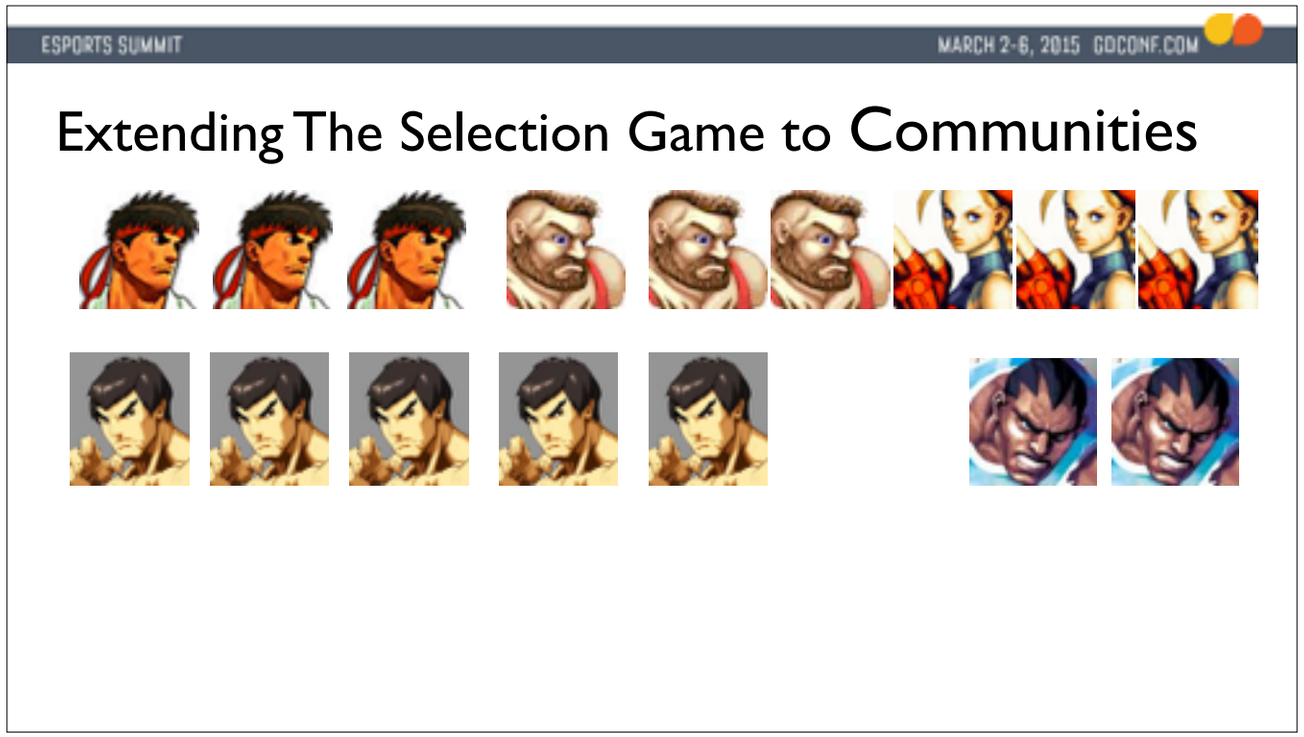
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



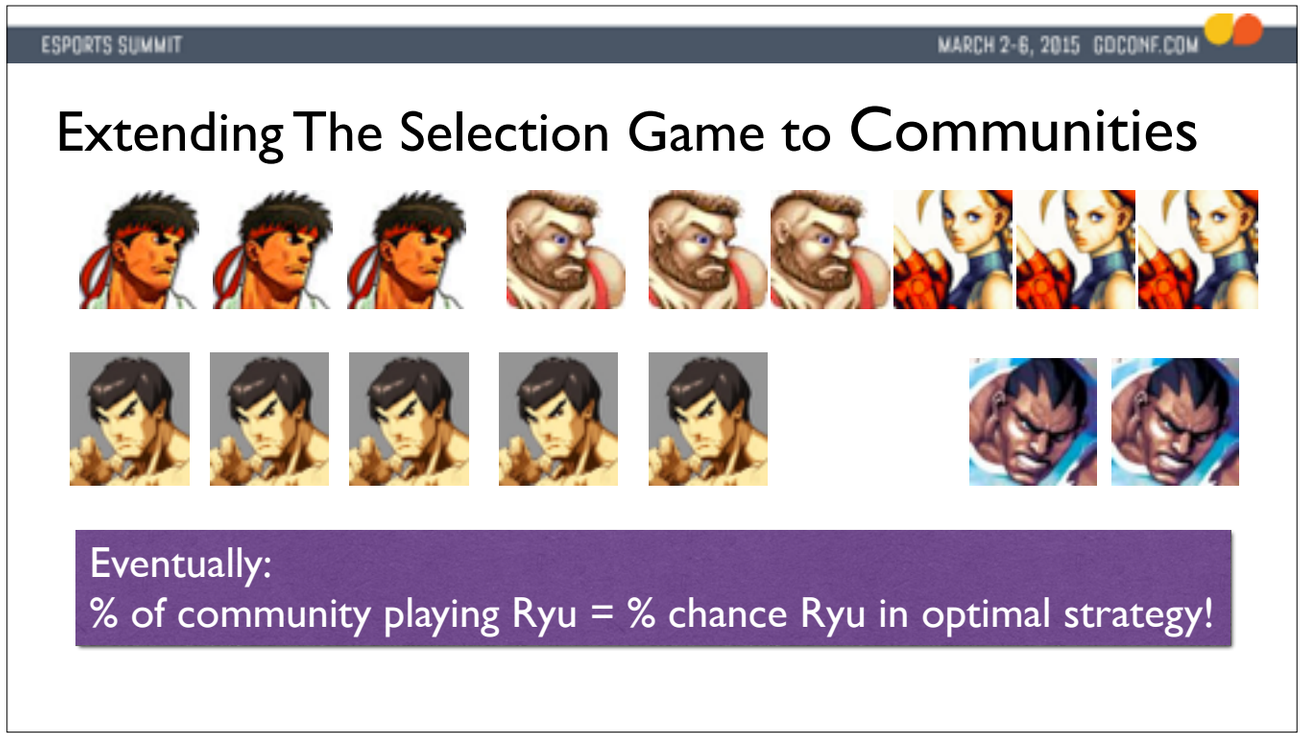
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



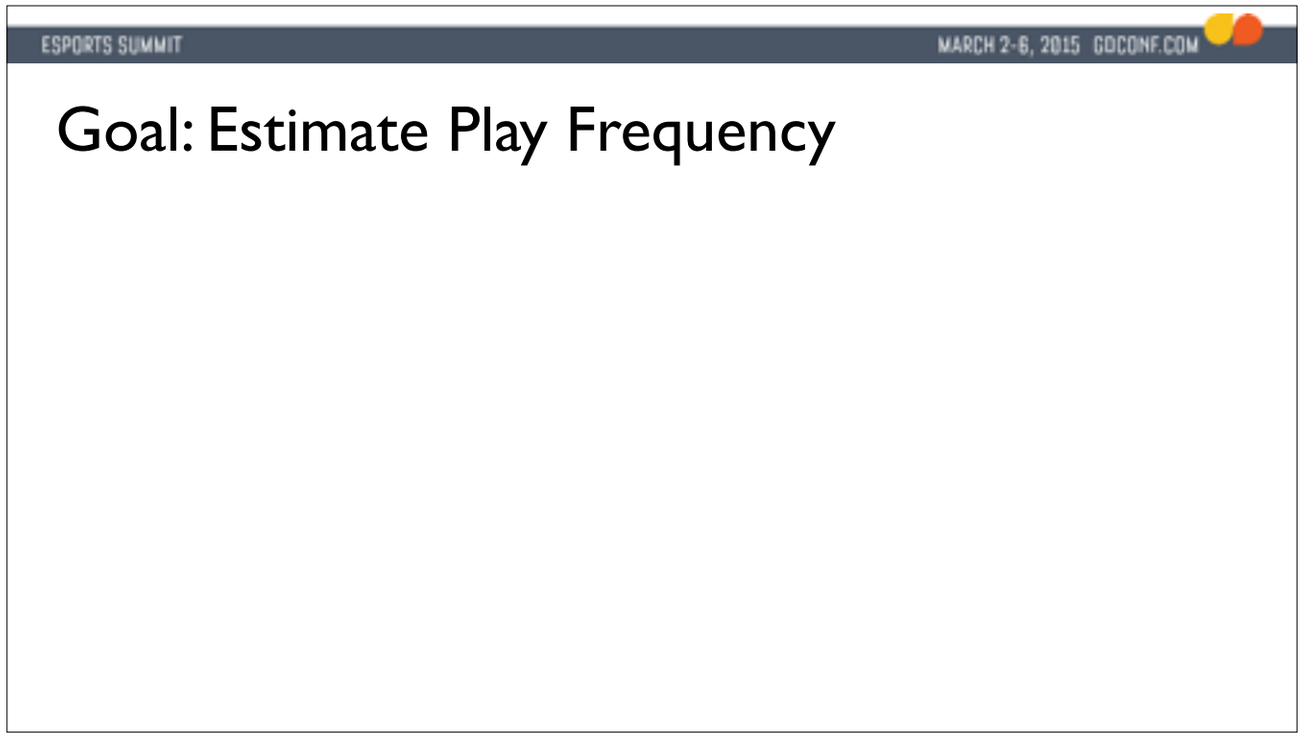
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



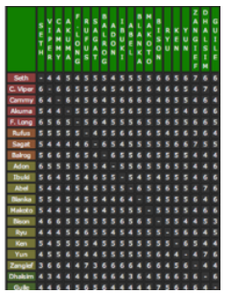
Maybe by this point some of you have the following complaint: in most games, players don't just pick an independent character each time. In the case of Street Fighter, they spend years training with one character, and are often quite hesitant to switch. A metagame is really a long-term, humanistic process that seems to push against the magic circle. I've got a surprising answer for that. Consider the following simplified model of how a community might form: each player sticks to a single character. Say you start out with 1/3 Ryu, 1/3 Zangief, and 1/3 Cammy, forming a natural light rock-paper-scissors. When a new player joins the game, they play for a while, then finally choose to stick with a character who performs the best against an average character in the current community. This player picks Fei Long, because he beats the original three. More players enter the game, also picking Fei Long, in order to dominate the existing community. Then eventually there are enough Fei Longs that someone picks Balrog, because of his advantage against Fei Long. So that's our very simplified model. But guess what? This process is identical to algorithm called "Fictitious Play", used to compute the optimal strategy of a Selection Game. That is, if the optimal strategy to SF4 says play Ryu 8% of the time, then over time the percentage of Ryu's in your community should approach 8%. Of course, that process isn't the real way players play a game, but for an execution-heavy game such as a fighting games, it has enough resemblance to reality to be heartening. Maybe these optimal strategies can say something meaningful about a game's community after all!



So now it seems we might be good. We should be able to take a game's matchups, and from it compute the optimal strategy. But remember that what we're looking for is a *range* of strategies, we want to know the max and min frequency a player might occur. If we do compute the optimal strategy we get this one. And it looks pretty off, if you know anything about Street Fighter. We could try searching for more optimal strategies, but in this case they're going to be pretty similar. Which gets to the next observation.

Our goal is both ambitious and modest. If we could estimate the frequency with which a given character is played that would be ridiculously powerful. And it certainly *seems* like we can do it. Previously we said that play frequency was too fragile, too subject to player whims and metagame changes. Yet that was empirical play frequency. Now we can talk about optimal play frequency as a measure of strength! It's optimal to play Balrog 32% of the time… But that's not right. Any SF4 player will tell you Balrog isn't the strongest character.

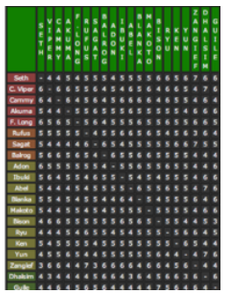
Goal: Estimate Play Frequency



So now it seems we might be good. We should be able to take a game's matchups, and from it compute the optimal strategy. But remember that what we're looking for is a *range* of strategies, we want to know the max and min frequency a player might occur. If we do compute the optimal strategy we get this one. And it looks pretty off, if you know anything about Street Fighter. We could try searching for more optimal strategies, but in this case they're going to be pretty similar. Which gets to the next observation.

Our goal is both ambitious and modest. If we could estimate the frequency with which a given character is played that would be ridiculously powerful. And it certainly *seems* like we can do it. Previously we said that play frequency was too fragile, too subject to player whims and metagame changes. Yet that was empirical play frequency. Now we can talk about optimal play frequency as a measure of strength! It's optimal to play Balrog 32% of the time… But that's not right. Any SF4 player will tell you Balrog isn't the strongest character.

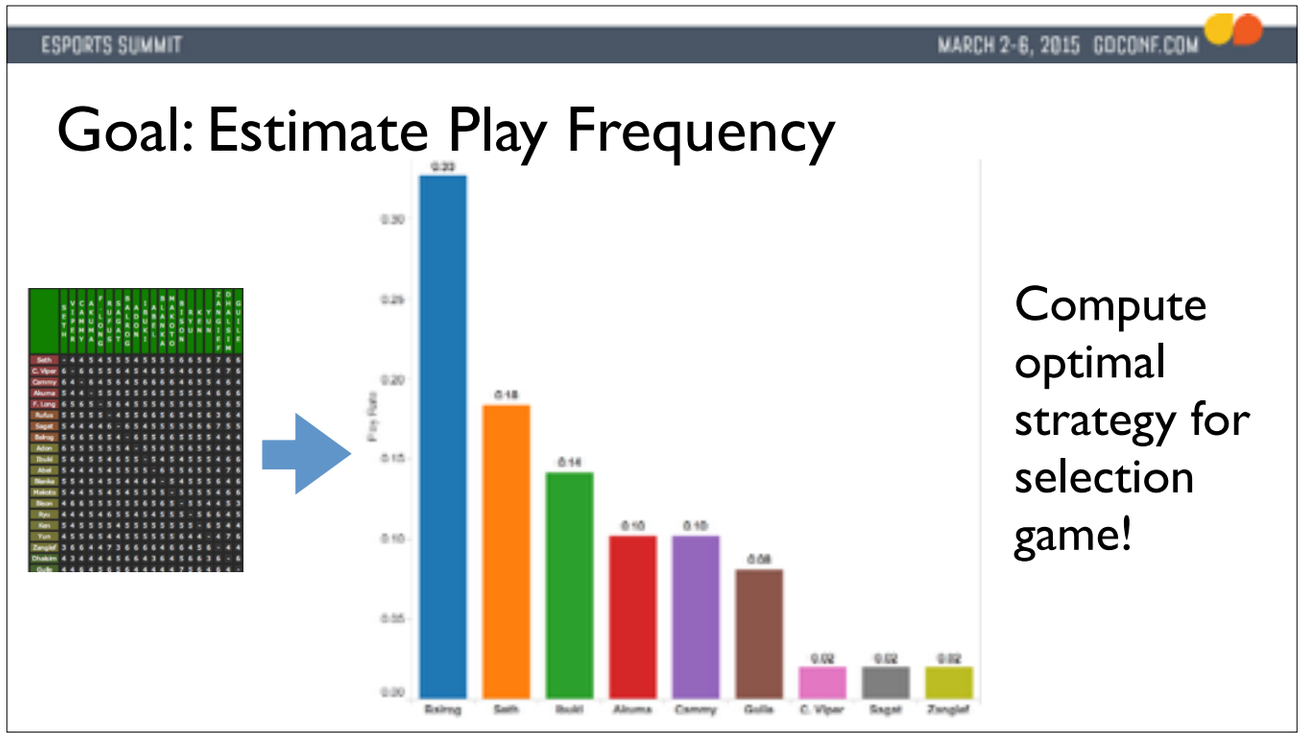
Goal: Estimate Play Frequency



Compute optimal strategy for selection game!

So now it seems we might be good. We should be able to take a game's matchups, and from it compute the optimal strategy. But remember that what we're looking for is a *range* of strategies, we want to know the max and min frequency a player might occur. If we do compute the optimal strategy we get this one. And it looks pretty off, if you know anything about Street Fighter. We could try searching for more optimal strategies, but in this case they're going to be pretty similar. Which gets to the next observation.

Our goal is both ambitious and modest. If we could estimate the frequency with which a given character is played that would be ridiculously powerful. And it certainly *seems* like we can do it. Previously we said that play frequency was too fragile, too subject to player whims and metagame changes. Yet that was empirical play frequency. Now we can talk about optimal play frequency as a measure of strength! It's optimal to play Balrog 32% of the time… But that's not right. Any SF4 player will tell you Balrog isn't the strongest character.



So now it seems we might be good. We should be able to take a game's matchups, and from it compute the optimal strategy. But remember that what we're looking for is a *range* of strategies, we want to know the max and min frequency a player might occur. If we do compute the optimal strategy we get this one. And it looks pretty off, if you know anything about Street Fighter. We could try searching for more optimal strategies, but in this case they're going to be pretty similar. Which gets to the next observation.

Our goal is both ambitious and modest. If we could estimate the frequency with which a given character is played that would be ridiculously powerful. And it certainly *seems* like we can do it. Previously we said that play frequency was too fragile, too subject to player whims and metagame changes. Yet that was empirical play frequency. Now we can talk about optimal play frequency as a measure of strength! It's optimal to play Balrog 32% of the time… But that's not right. Any SF4 player will tell you Balrog isn't the strongest character.

Perfectly Optimal isn't Realistic

The problem is that perfectly optimal play isn't actually realistic! Players play well, but they don't play perfectly, and what's more the matchups are noisy, even if they're from real data. So the big idea here is to also look at approximately optimal strategies. We *force* the solver to play some character some amount, then we see how well it does. And this is as simple as adding one constraint to the linear program. We ask "can we win almost half the games with 12% Balrog? With 13% Balrog? With 18% Blanka?" By asking hundreds of such questions, we get much more diverse mixed strategies, and essentially create a map of the ranges of frequencies at which a character can be played in communities that can't be dominated by an advantaged character.

Perfectly Optimal isn't Realistic

How about approximately optimal? Force strategy to play Balrog 12% of games. Can we still win, say 45% of games?

The problem is that perfectly optimal play isn't actually realistic! Players play well, but they don't play perfectly, and what's more the matchups are noisy, even if they're from real data. So the big idea here is to also look at approximately optimal strategies. We *force* the solver to play some character some amount, then we see how well it does. And this is as simple as adding one constraint to the linear program. We ask "can we win almost half the games with 12% Balrog? With 13% Balrog? With 18% Blanka?" By asking hundreds of such questions, we get much more diverse mixed strategies, and essentially create a map of the ranges of frequencies at which a character can be played in communities that can't be dominated by an advantaged character.

Perfectly Optimal isn't Realistic

Modified
Linear Program

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{im}x_i\right)\right\}$$

$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

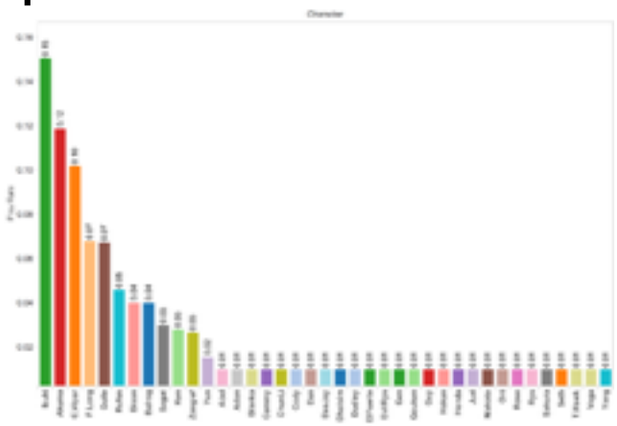
$$x_i \geq 0 \qquad \forall i$$

x_i = 0.12

How about approximately optimal? Force strategy to play Balrog 12% of games. Can we still win, say 45% of games?

The problem is that perfectly optimal play isn't actually realistic! Players play well, but they don't play perfectly, and what's more the matchups are noisy, even if they're from real data. So the big idea here is to also look at approximately optimal strategies. We *force* the solver to play some character some amount, then we see how well it does. And this is as simple as adding one constraint to the linear program. We ask "can we win almost half the games with 12% Balrog? With 13% Balrog? With 18% Blanka?" By asking hundreds of such questions, we get much more diverse mixed strategies, and essentially create a map of the ranges of frequencies at which a character can be played in communities that can't be dominated by an advantaged character.

Perfectly Optimal isn't Realistic



Modified
Linear Program

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{im}x_i\right)\right\}$$

$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

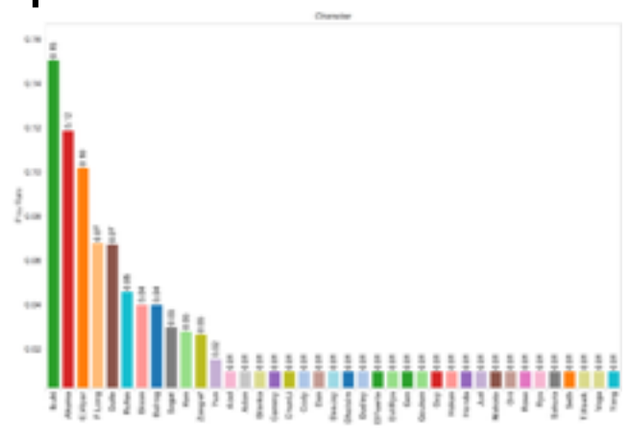
$$x_i \geq 0 \qquad \forall i$$

x_i = 0.12

How about approximately optimal? Force strategy to play Balrog 12% of games. Can we still win, say 45% of games?

The problem is that perfectly optimal play isn't actually realistic! Players play well, but they don't play perfectly, and what's more the matchups are noisy, even if they're from real data. So the big idea here is to also look at approximately optimal strategies. We *force* the solver to play some character some amount, then we see how well it does. And this is as simple as adding one constraint to the linear program. We ask "can we win almost half the games with 12% Balrog? With 13% Balrog? With 18% Blanka?" By asking hundreds of such questions, we get much more diverse mixed strategies, and essentially create a map of the ranges of frequencies at which a character can be played in communities that can't be dominated by an advantaged character.

Perfectly Optimal isn't Realistic



Modified
Linear Program

$$\max_{x}\left\{\min\left(\sum_{i=1}^{m} v_{i1}x_i, \sum_{i=1}^{m} v_{i2}x_i, \ldots, \sum_{i=1}^{m} v_{in}x_i\right)\right\}$$

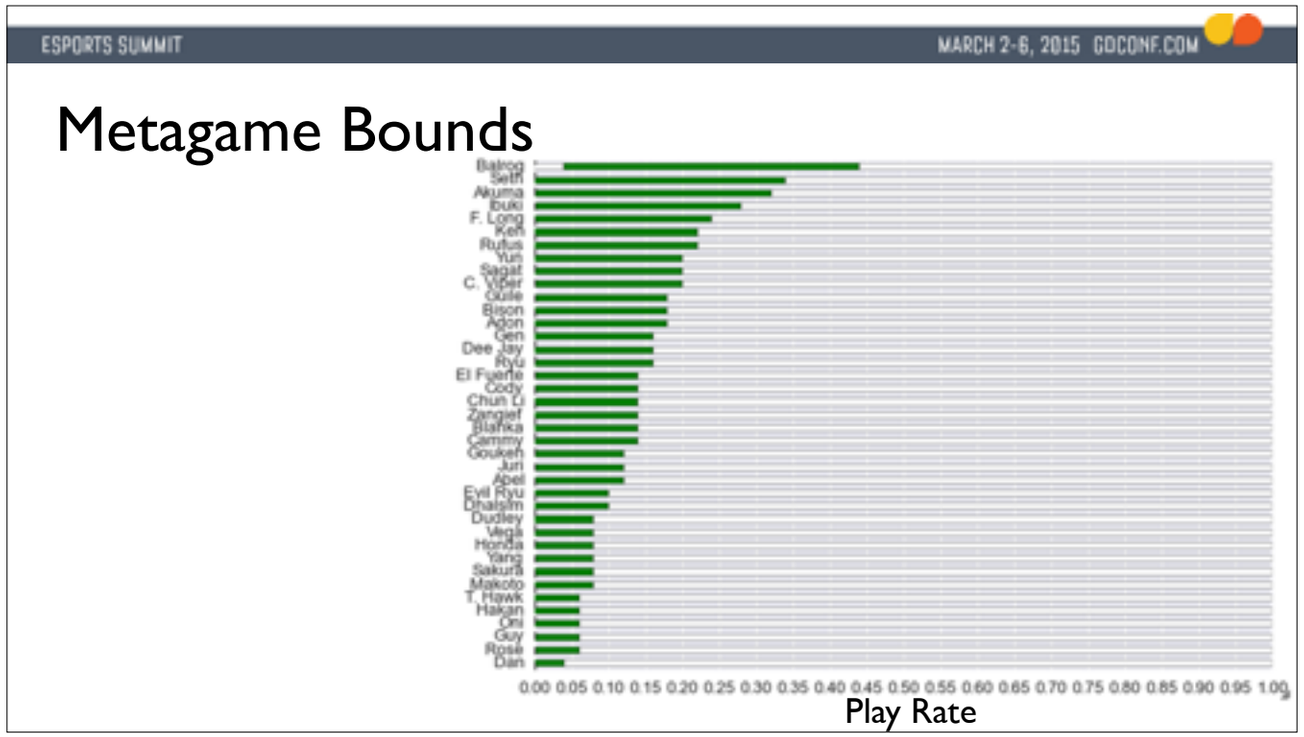
$$s.t. \quad \sum_{i=1}^{m} x_i = 1$$

$$x_i \geq 0 \qquad \forall i$$

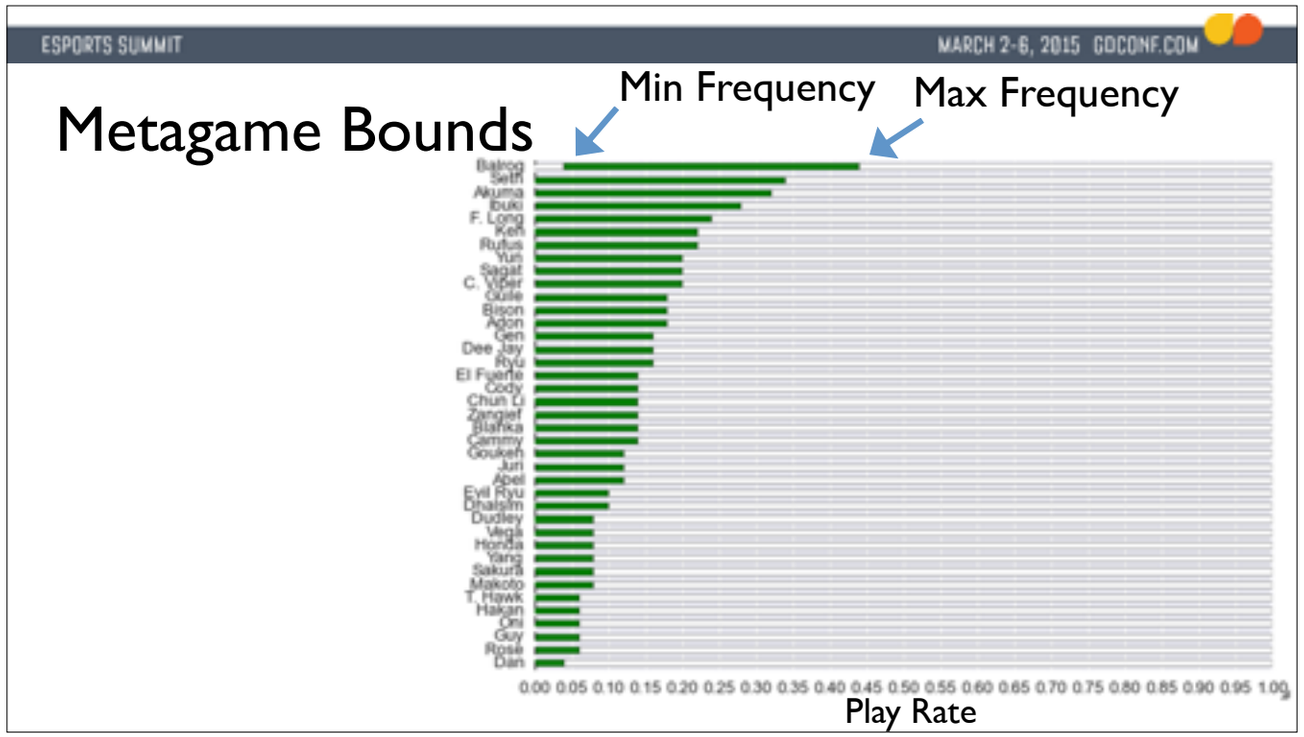
$$x_i = 0.12$$

How about approximately optimal? Force strategy to play Balrog 12% of games. Can we still win, say 45% of games?

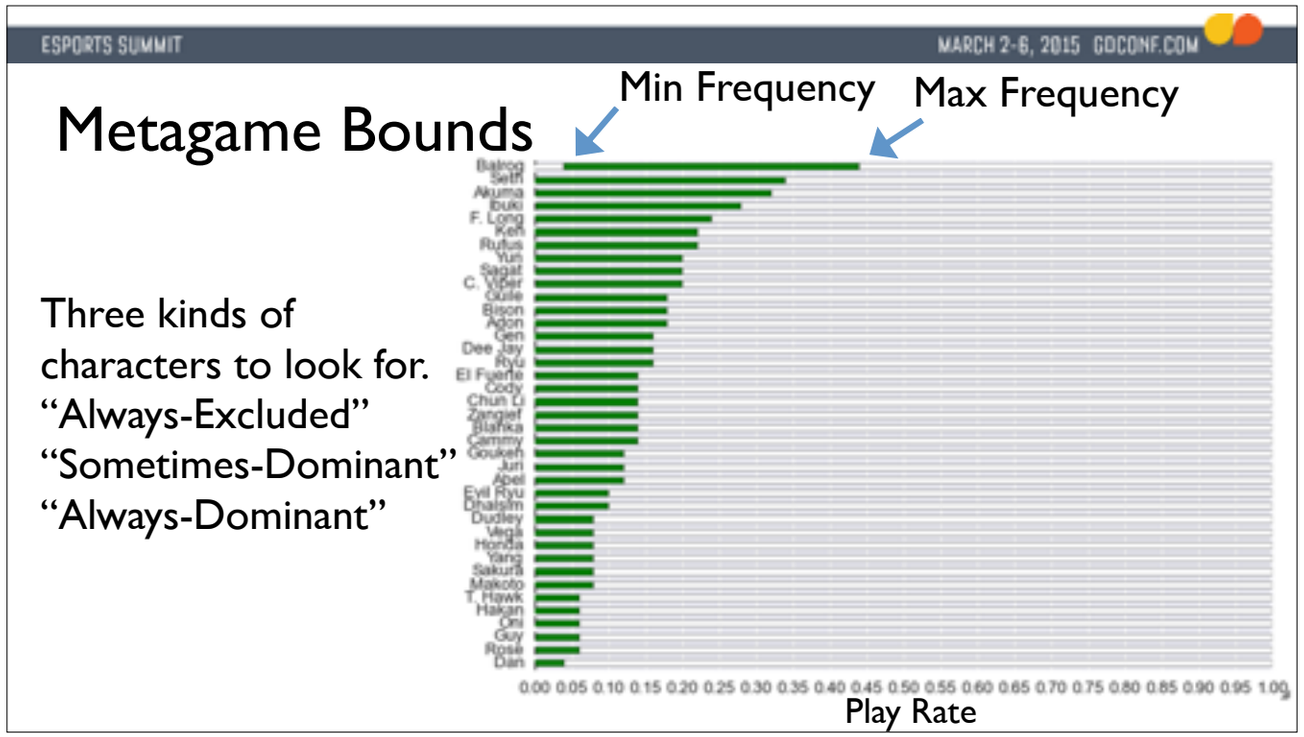
The problem is that perfectly optimal play isn't actually realistic! Players play well, but they don't play perfectly, and what's more the matchups are noisy, even if they're from real data. So the big idea here is to also look at approximately optimal strategies. We *force* the solver to play some character some amount, then we see how well it does. And this is as simple as adding one constraint to the linear program. We ask "can we win almost half the games with 12% Balrog? With 13% Balrog? With 18% Blanka?" By asking hundreds of such questions, we get much more diverse mixed strategies, and essentially create a map of the ranges of frequencies at which a character can be played in communities that can't be dominated by an advantaged character.



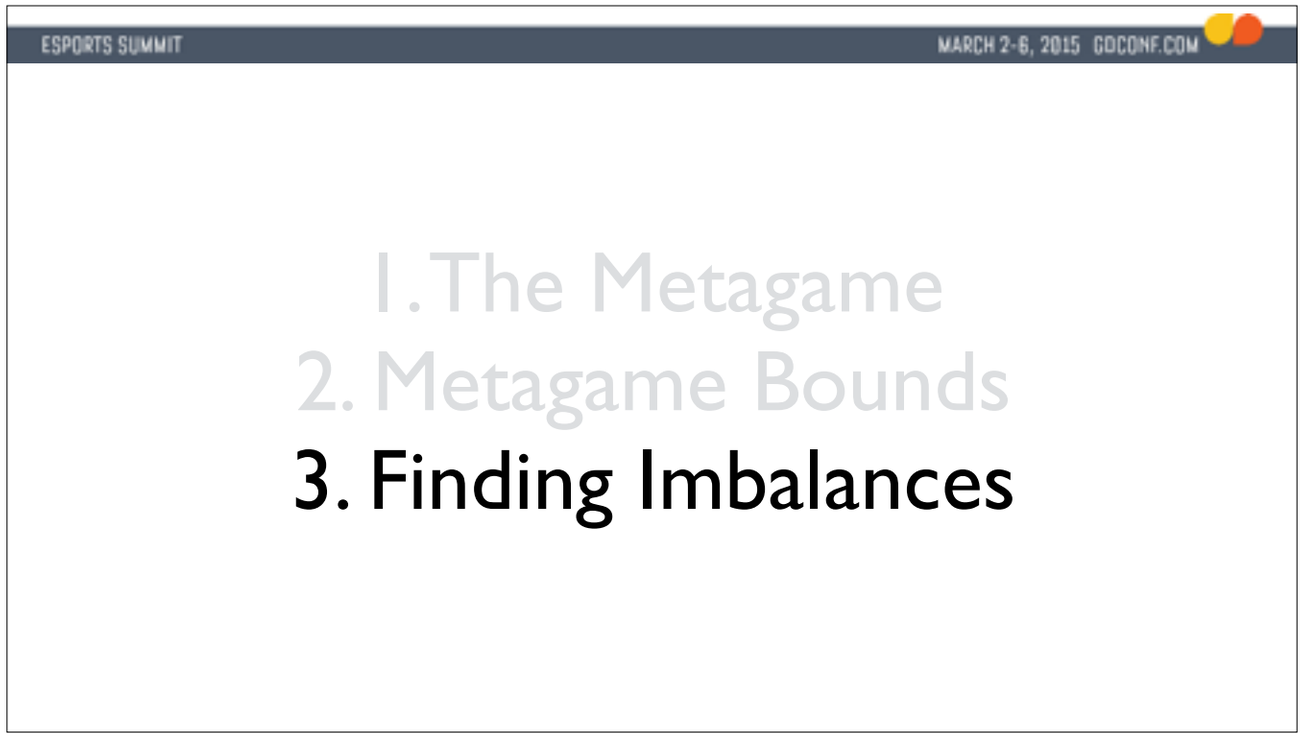
And this gets what we want! We draw a green interval stretching from the minimum “valid” frequency to the maximum “valid” frequency (included in a mixed strategy that wins at least 45%, say, of the time). We get a map of many possible communites, almost alternate realities. Here Balrog must be played at least 4% of the time, and at most 44% of the time. Any character can be played at least a little bit, and no one has to be played a lot. This is the the final product of this method, though just one of several ways I’ve explored analyzing matchup charts. By looking at these graphs, we can answer several kinds of questions where the metagame might go. Three, for example. “Always-Dominant Characters”, “Sometimes-Dominant Characters”, and “Always-Excluded Characters”.



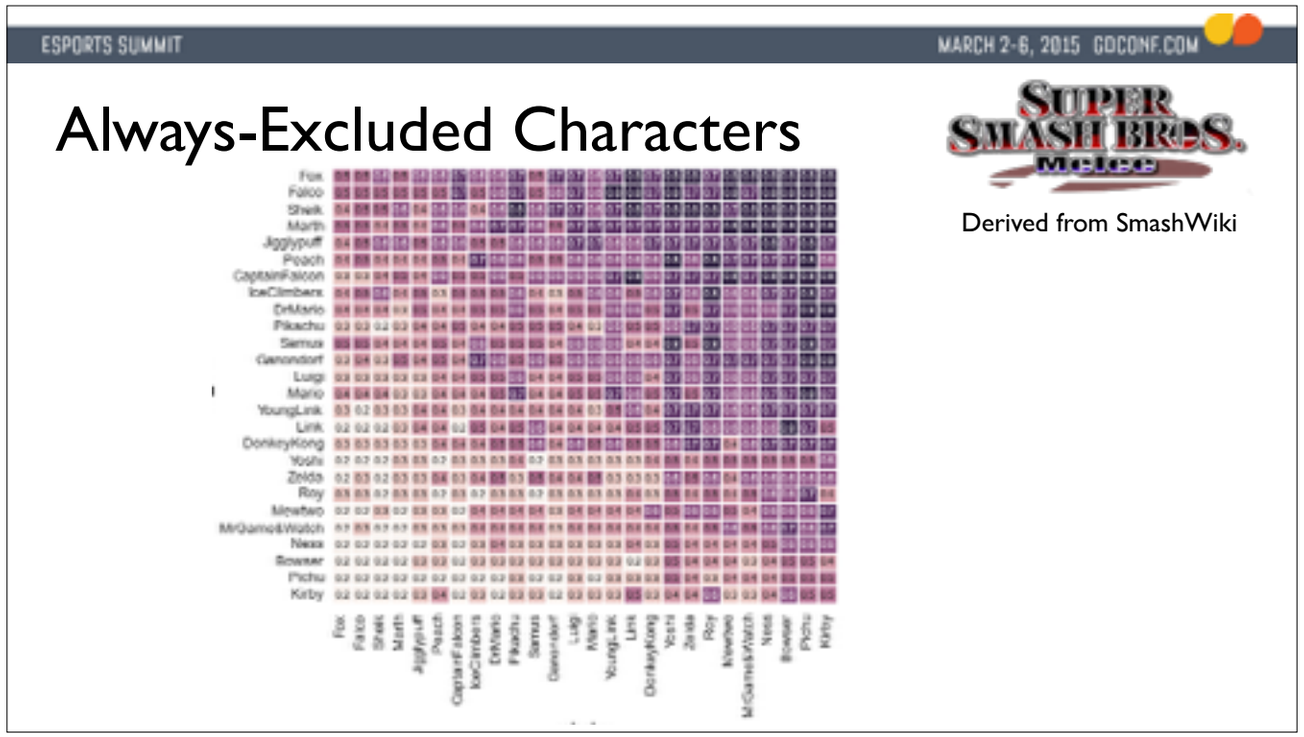
And this gets what we want! We draw a green interval stretching from the minimum “valid” frequency to the maximum “valid” frequency (included in a mixed strategy that wins at least 45%, say, of the time). We get a map of many possible communites, almost alternate realities. Here Balrog must be played at least 4% of the time, and at most 44% of the time. Any character can be played at least a little bit, and no one has to be played a lot. This is the the final product of this method, though just one of several ways I’ve explored analyzing matchup charts. By looking at these graphs, we can answer several kinds of questions where the metagame might go. Three, for example. “Always-Dominant Characters”, “Sometimes-Dominant Characters”, and “Always-Excluded Characters”.



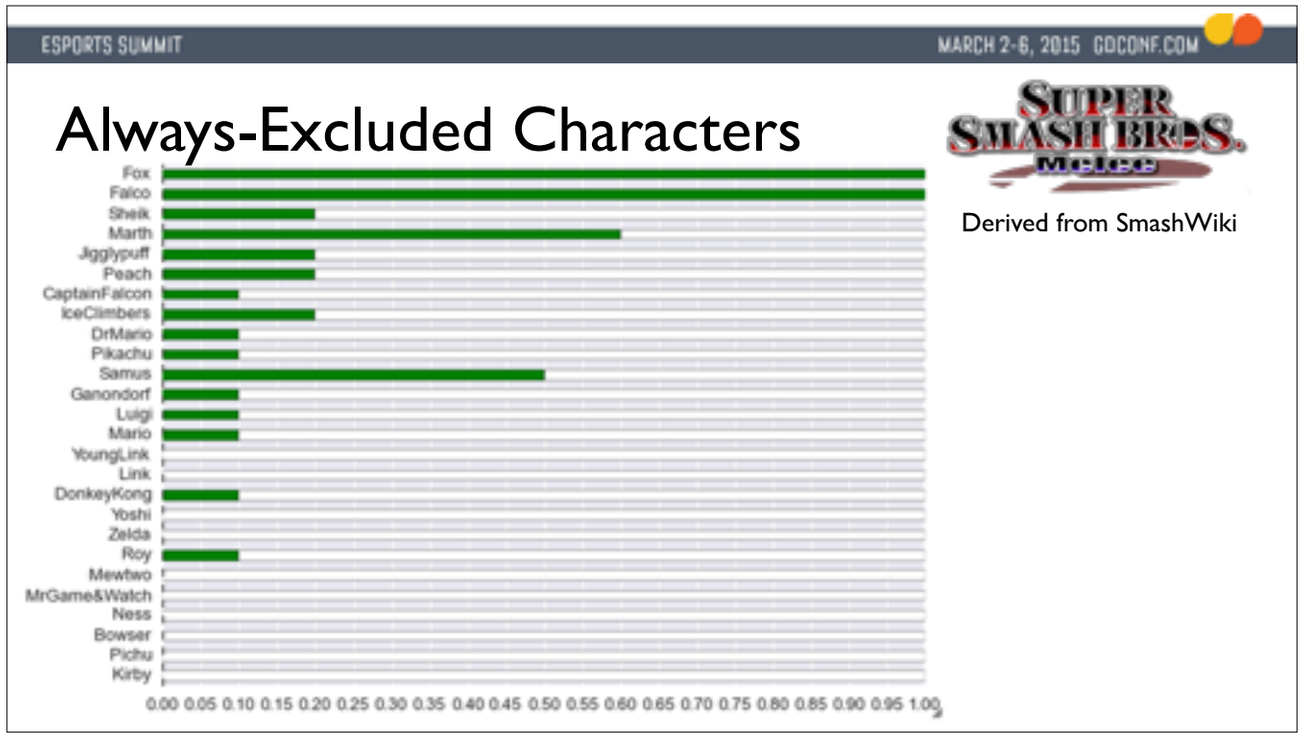
And this gets what we want! We draw a green interval stretching from the minimum “valid” frequency to the maximum “valid” frequency (included in a mixed strategy that wins at least 45%, say, of the time). We get a map of many possible communites, almost alternate realities. Here Balrog must be played at least 4% of the time, and at most 44% of the time. Any character can be played at least a little bit, and no one has to be played a lot. This is the the final product of this method, though just one of several ways I’ve explored analyzing matchup charts. By looking at these graphs, we can answer several kinds of questions where the metagame might go. Three, for example. “Always-Dominant Characters”, “Sometimes-Dominant Characters”, and “Always-Excluded Characters”.



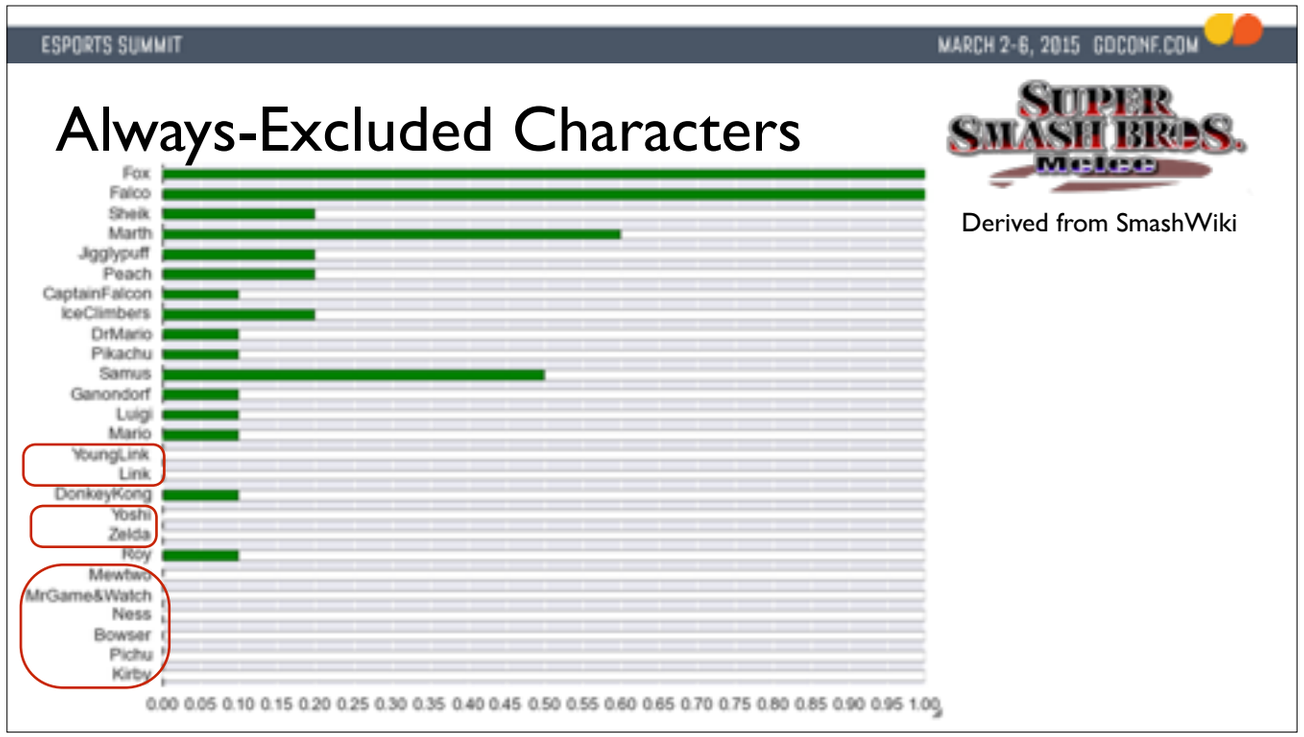
I'm now going to give a few examples of creating and reading metagame bounds for estimated data from a few real games.



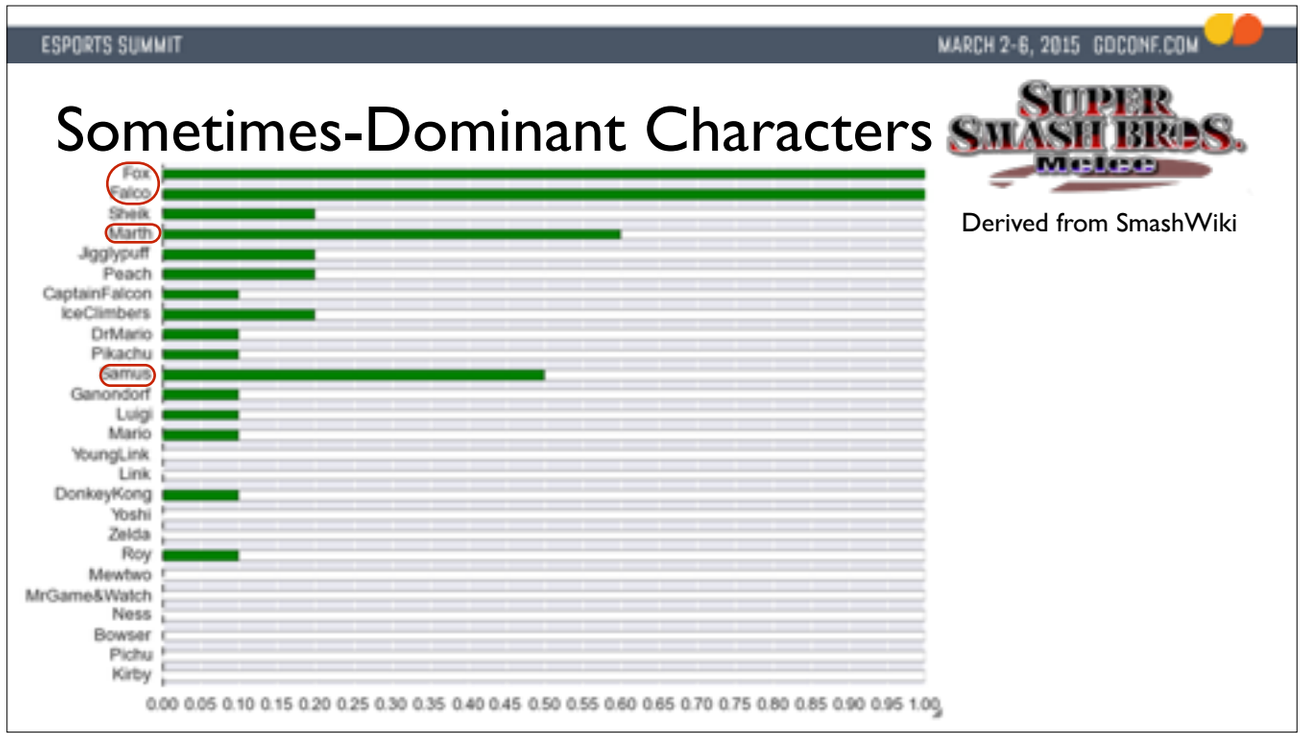
We haven't talked much about underpowered characters, but it is a concern. Consider this chart of matchup data, derived from some player-created charts on SmashWiki. We can run it through the algorithm I just described, and it outputs these interval graphs. They give you a pretty easy way to detect characters who are underpowered in every competitive metagame. Young Link, Yoshi, Mewto, etc. shouldn't be expected to appear basically at all. Player freedom is a major goal: letting them express themselves through play. If players can't pick a certain character without getting destroyed, in every strong community, that's often a problem! With just a few matchup changes, you can at least create the potential for a character to be useful.



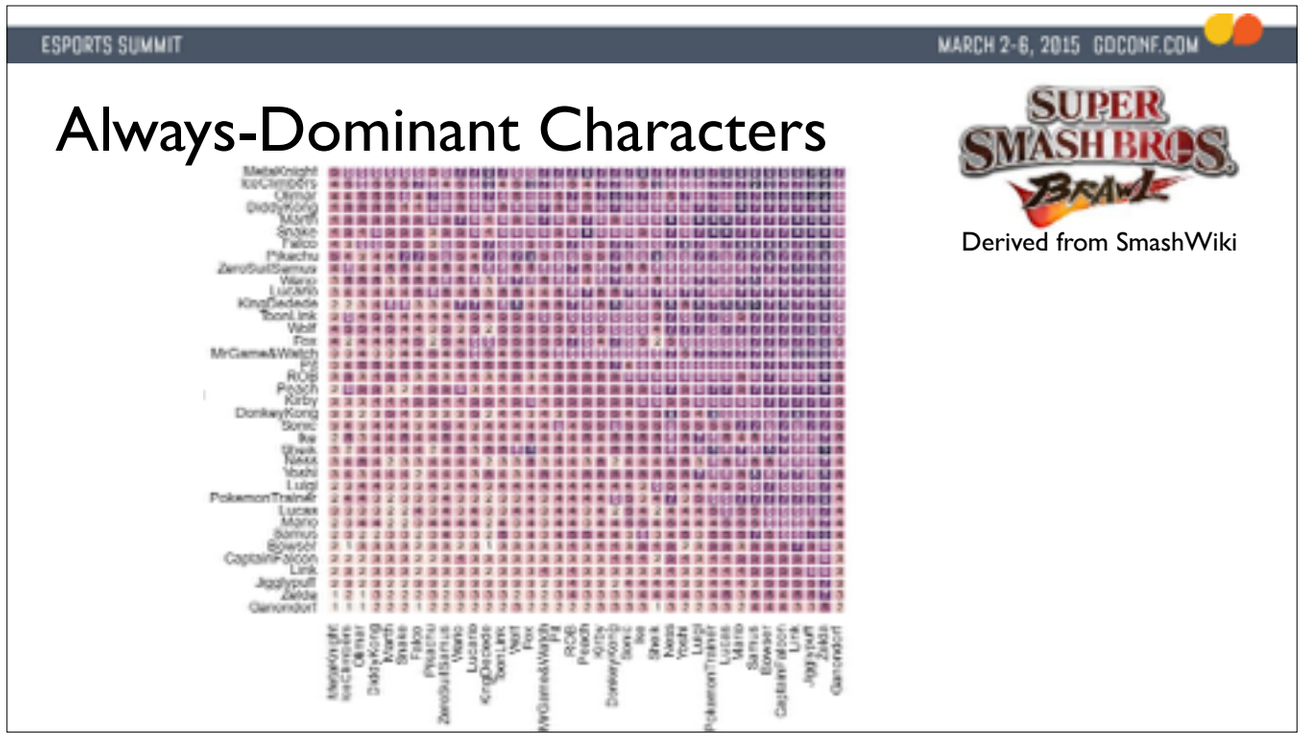
We haven't talked much about underpowered characters, but it is a concern. Consider this chart of matchup data, derived from some player-created charts on SmashWiki. We can run it through the algorithm I just described, and it outputs these interval graphs. They give you a pretty easy way to detect characters who are underpowered in every competitive metagame. Young Link, Yoshi, Mewto, etc. shouldn't be expected to appear basically at all. Player freedom is a major goal: letting them express themselves through play. If players can't pick a certain character without getting destroyed, in every strong community, that's often a problem! With just a few matchup changes, you can at least create the potential for a character to be useful.



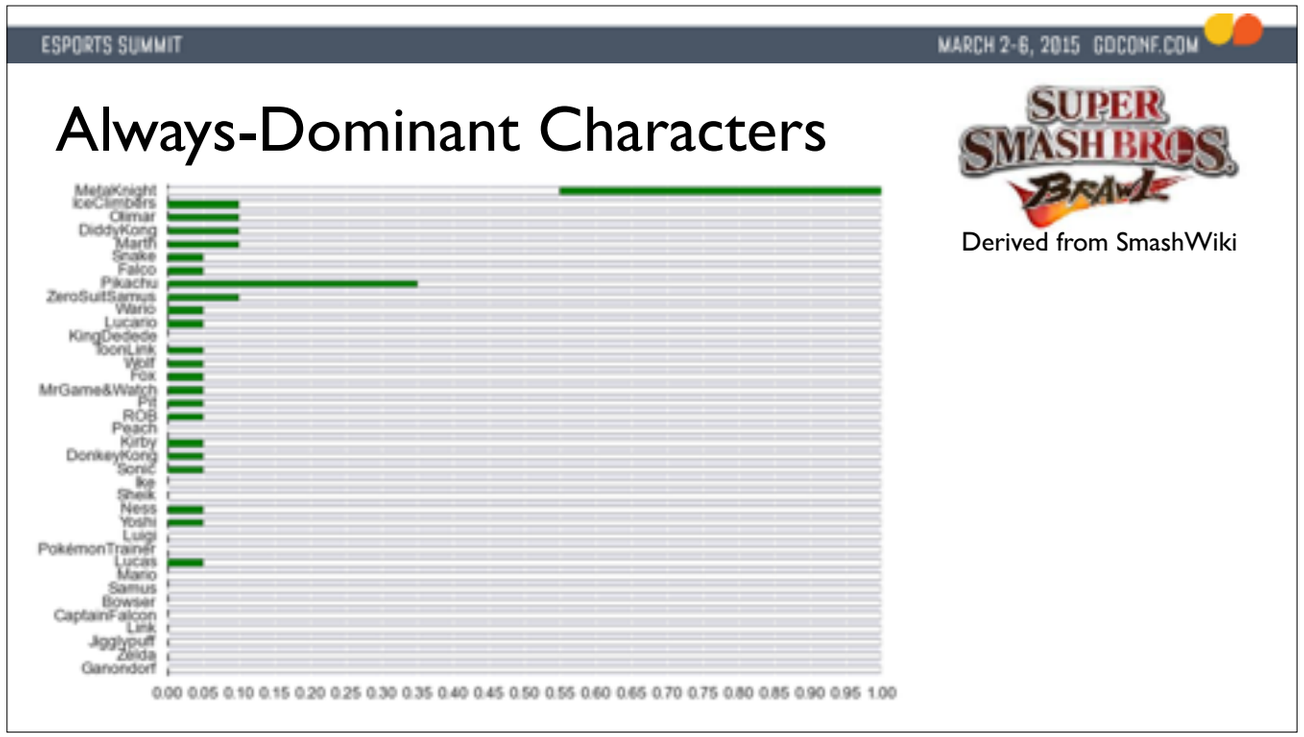
We haven't talked much about underpowered characters, but it is a concern. Consider this chart of matchup data, derived from some player-created charts on SmashWiki. We can run it through the algorithm I just described, and it outputs these interval graphs. They give you a pretty easy way to detect characters who are underpowered in every competitive metagame. Young Link, Yoshi, Mewto, etc. shouldn't be expected to appear basically at all. Player freedom is a major goal: letting them express themselves through play. If players can't pick a certain character without getting destroyed, in every strong community, that's often a problem! With just a few matchup changes, you can at least create the potential for a character to be useful.



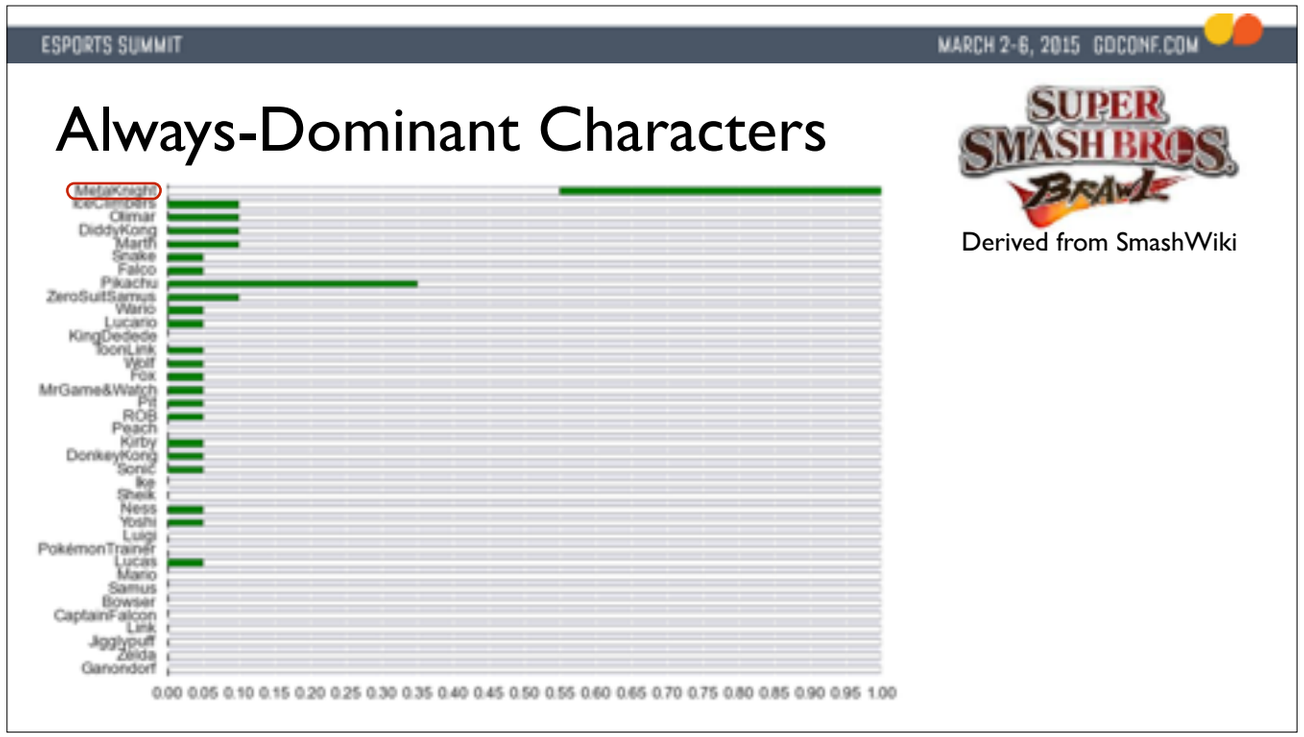
Next up: characters who can be played frequently *or* can be played less. Here it's possible that this character could overwhelm your community, but it's hard to tell in advance, since it depends on the whims of your players. Is this a problem? Depends! It would be unpleasant if 90% of players played one character, for players and especially for spectators. This kind of thing isn't common - players self-distribute. And yet we did see some of this in All-Stars. Players have natural preferences, e.g. for Kratos, and they will play him ad nauseam, provided it's a viable option. This is largely a matter of designer taste. Maybe you're okay with this. But what can you do if you're worried about players clumping around one or two characters, and want to prevent it? It comes at a cost. You want to find ways to selectively give a character bad matchups against other good characters. (E.g. giving an opponent a soft counter against them.) Note they can be given other good matchups to compensate. Maybe you don't want any biased matchups in your game, and that's your choice. But I argue, as Seth Killian said to me, that biased matchups aren't that bad, provided the disadvantaged character has a clear path to winning, however unlikely it might be.



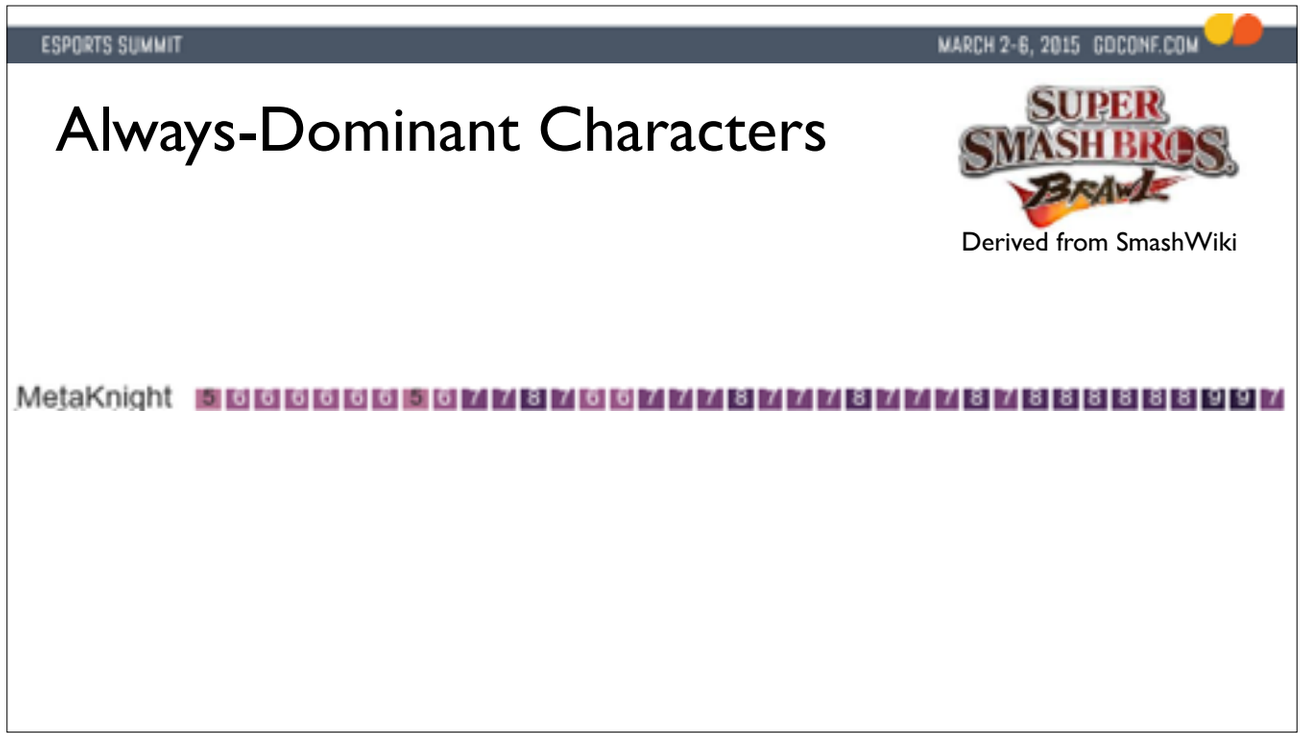
Perhaps the biggest danger to look out for are characters for whom *any* near optimal strategy features a huge percentage of that character. Here's another matchup chart, this time for Smash Bros. Brawl. Those familiar with the game will know that it has such a character in Metaknight. Plugging the matchup charts in produces the result you'd expect. You can't get near optimal without playing him at least 60% of the time! (I.e. a community that is resistant to outside aggression has at least 60% Metaknights.) Yet if you're looking closely, you'd see that this shouldn't be surprising. Metaknight's matchups are far and away stronger than any other character's. *Of course* he dominates, right?



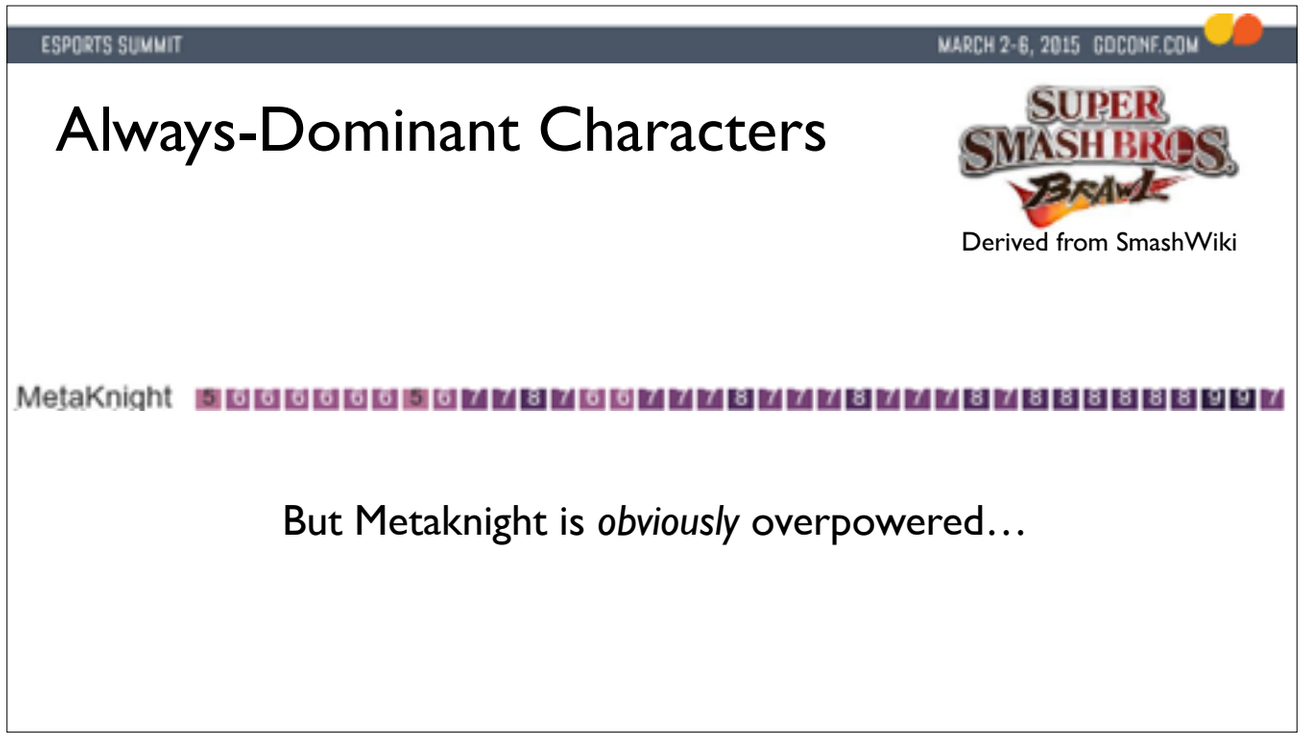
Perhaps the biggest danger to look out for are characters for whom *any* near optimal strategy features a huge percentage of that character. Here's another matchup chart, this time for Smash Bros. Brawl. Those familiar with the game will know that it has such a character in Metaknight. Plugging the matchup charts in produces the result you'd expect. You can't get near optimal without playing him at least 60% of the time! (I.e. a community that is resistant to outside aggression has at least 60% Metaknights.) Yet if you're looking closely, you'd see that this shouldn't be surprising. Metaknight's matchups are far and away stronger than any other character's. *Of course* he dominates, right?



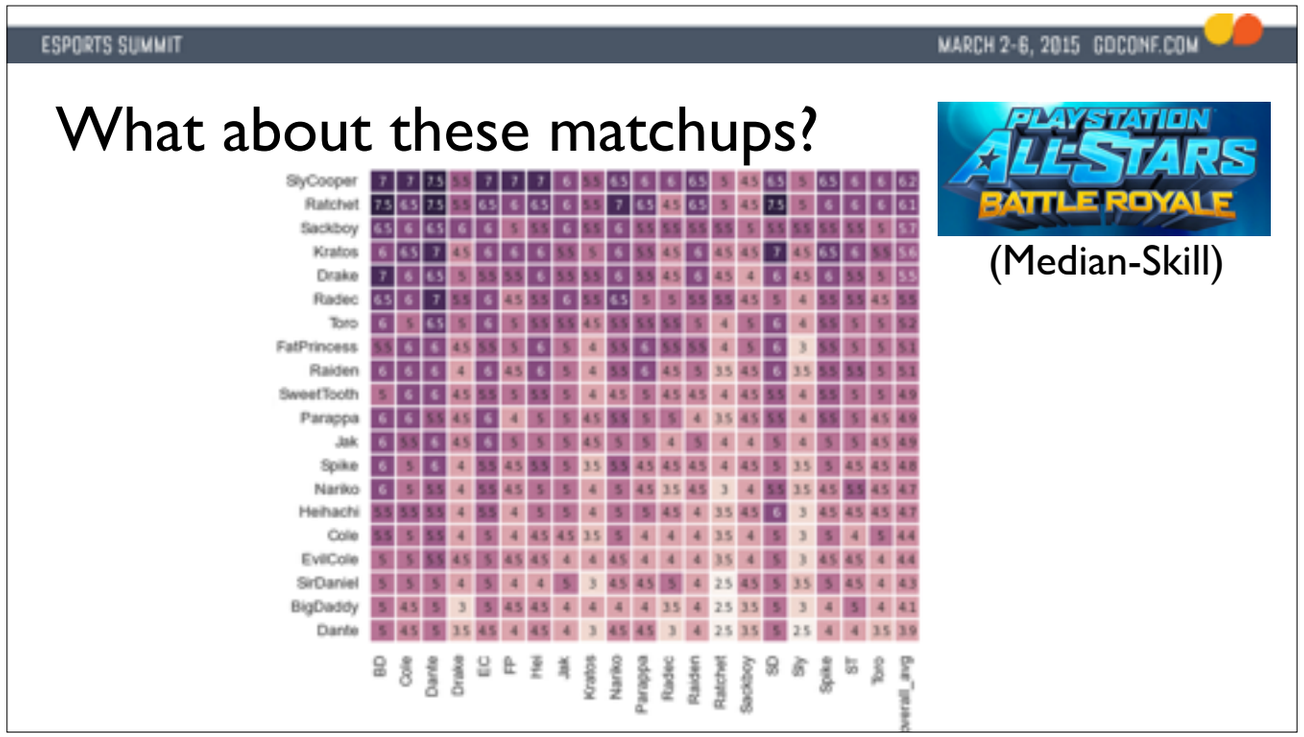
Perhaps the biggest danger to look out for are characters for whom *any* near optimal strategy features a huge percentage of that character. Here's another matchup chart, this time for Smash Bros. Brawl. Those familiar with the game will know that it has such a character in Metaknight. Plugging the matchup charts in produces the result you'd expect. You can't get near optimal without playing him at least 60% of the time! (I.e. a community that is resistant to outside aggression has at least 60% Metaknights.) Yet if you're looking closely, you'd see that this shouldn't be surprising. Metaknight's matchups are far and away stronger than any other character's. *Of course* he dominates, right?



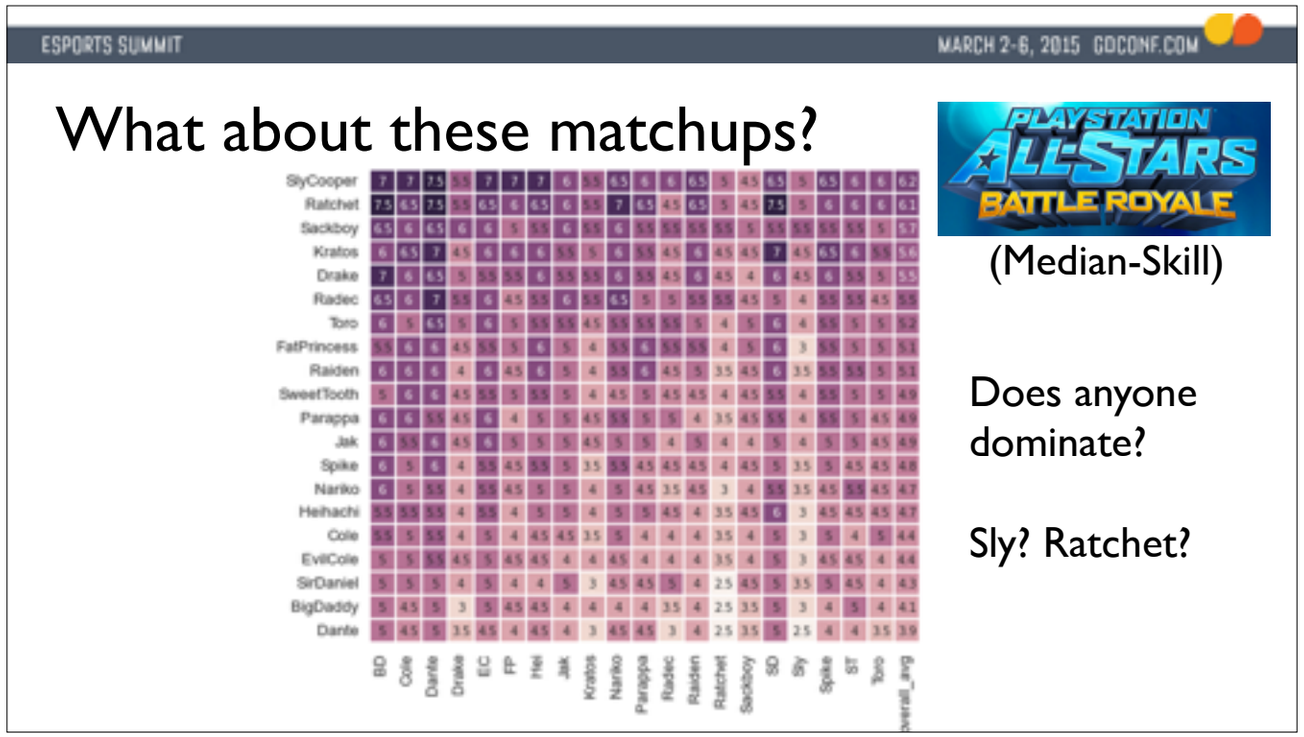
Perhaps the biggest danger to look out for are characters for whom *any* near optimal strategy features a huge percentage of that character. Here's another matchup chart, this time for Smash Bros. Brawl. Those familiar with the game will know that it has such a character in Metaknight. Plugging the matchup charts in produces the result you'd expect. You can't get near optimal without playing him at least 60% of the time! (I.e. a community that is resistant to outside aggression has at least 60% Metaknights.) Yet if you're looking closely, you'd see that this shouldn't be surprising. Metaknight's matchups are far and away stronger than any other character's. *Of course* he dominates, right?



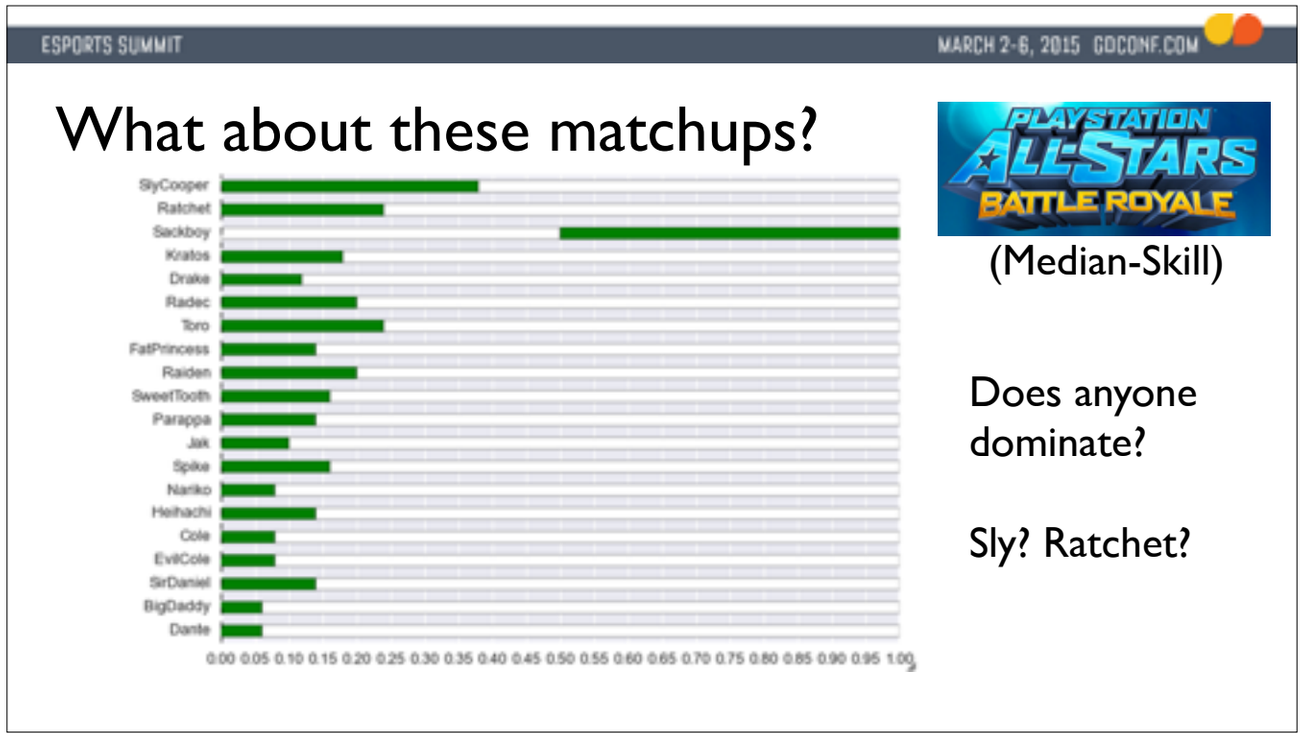
Perhaps the biggest danger to look out for are characters for whom *any* near optimal strategy features a huge percentage of that character. Here's another matchup chart, this time for Smash Bros. Brawl. Those familiar with the game will know that it has such a character in Metaknight. Plugging the matchup charts in produces the result you'd expect. You can't get near optimal without playing him at least 60% of the time! (I.e. a community that is resistant to outside aggression has at least 60% Metaknights.) Yet if you're looking closely, you'd see that this shouldn't be surprising. Metaknight's matchups are far and away stronger than any other character's. *Of course* he dominates, right?



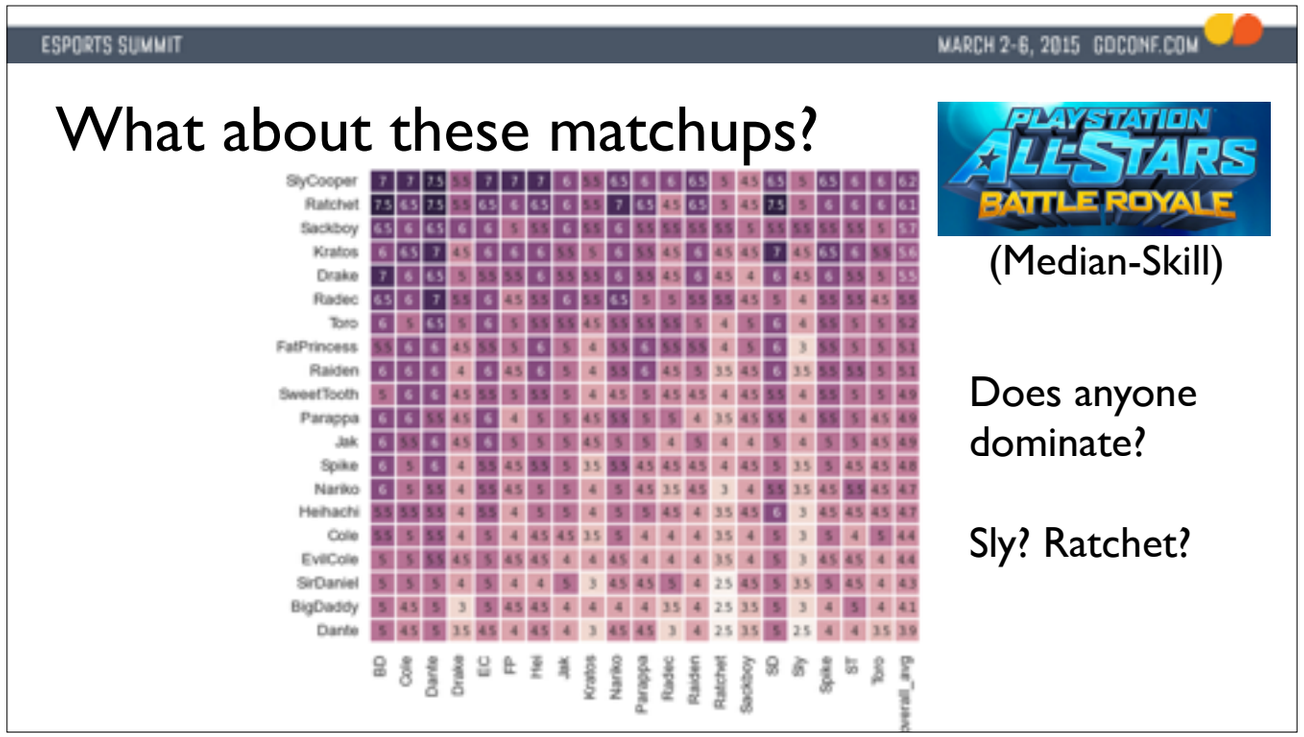
So consider this data instead. Here are some matchups from the first few weeks of PlayStation All-Stars, even before the patch from which those IGN tier lists at were created. The matchups are inferred from 4-player data, through a great deal of effort, but still mostly right. It's covering median-skill players b/c you need extra data to infer matchups from Free-For-All games, and they have the most data. From these matchups, can you tell if anyone dominates? There's a pretty smooth fall-off of average matchups. And yet, when we plug the numbers into the algorithm, we get a situation very much like that for Smash Brawl. Sackboy completely dominates the metagame. Every strong community has at least 50% of him. And that is very hard to see from the matchups! At the high level it's because he has no bad matchups, and has good matchups against otherwise strong characters. But there's a lot of nuanced interactions. Each character exerts a pressure on all others, there are cyclic relationships that are hard to reason about, but this method lays it bare for you, showing approximately how the whole thing might pan out.



So consider this data instead. Here are some matchups from the first few weeks of PlayStation All-Stars, even before the patch from which those IGN tier lists at were created. The matchups are inferred from 4-player data, through a great deal of effort, but still mostly right. It's covering median-skill players b/c you need extra data to infer matchups from Free-For-All games, and they have the most data. From these matchups, can you tell if anyone dominates? There's a pretty smooth fall-off of average matchups. And yet, when we plug the numbers into the algorithm, we get a situation very much like that for Smash Brawl. Sackboy completely dominates the metagame. Every strong community has at least 50% of him. And that is very hard to see from the matchups! At the high level it's because he has no bad matchups, and has good matchups against otherwise strong characters. But there's a lot of nuanced interactions. Each character exerts a pressure on all others, there are cyclic relationships that are hard to reason about, but this method lays it bare for you, showing approximately how the whole thing might pan out.



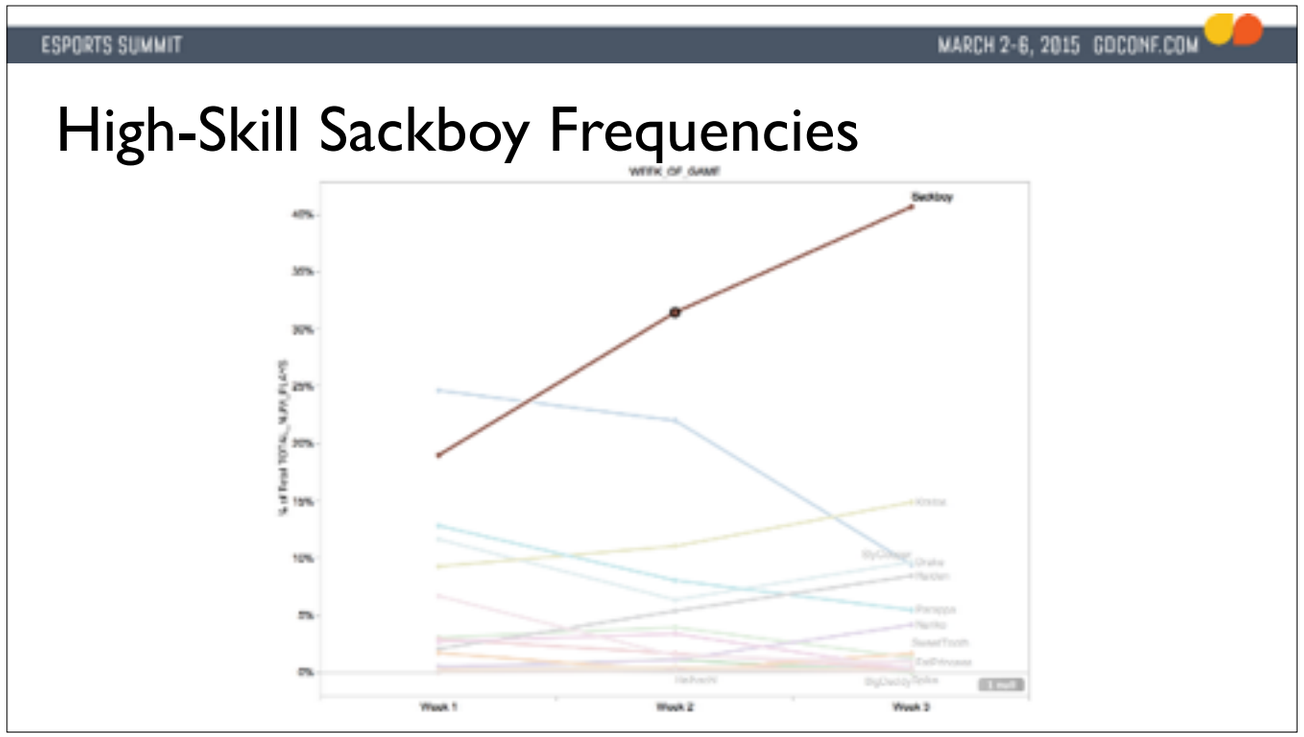
So consider this data instead. Here are some matchups from the first few weeks of PlayStation All-Stars, even before the patch from which those IGN tier lists at were created. The matchups are inferred from 4-player data, through a great deal of effort, but still mostly right. It's covering median-skill players b/c you need extra data to infer matchups from Free-For-All games, and they have the most data. From these matchups, can you tell if anyone dominates? There's a pretty smooth fall-off of average matchups. And yet, when we plug the numbers into the algorithm, we get a situation very much like that for Smash Brawl. Sackboy completely dominates the metagame. Every strong community has at least 50% of him. And that is very hard to see from the matchups! At the high level it's because he has no bad matchups, and has good matchups against otherwise strong characters. But there's a lot of nuanced interactions. Each character exerts a pressure on all others, there are cyclic relationships that are hard to reason about, but this method lays it bare for you, showing approximately how the whole thing might pan out.



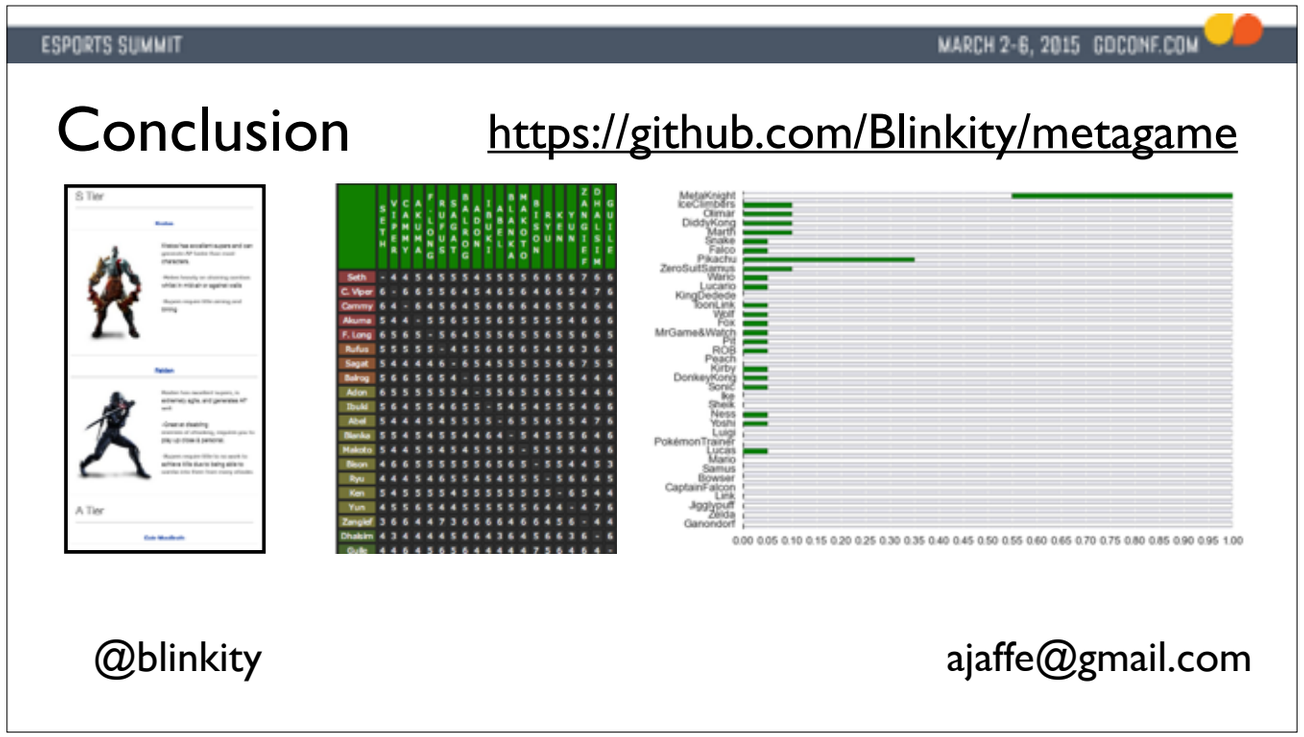
So consider this data instead. Here are some matchups from the first few weeks of PlayStation All-Stars, even before the patch from which those IGN tier lists at were created. The matchups are inferred from 4-player data, through a great deal of effort, but still mostly right. It's covering median-skill players b/c you need extra data to infer matchups from Free-For-All games, and they have the most data. From these matchups, can you tell if anyone dominates? There's a pretty smooth fall-off of average matchups. And yet, when we plug the numbers into the algorithm, we get a situation very much like that for Smash Brawl. Sackboy completely dominates the metagame. Every strong community has at least 50% of him. And that is very hard to see from the matchups! At the high level it's because he has no bad matchups, and has good matchups against otherwise strong characters. But there's a lot of nuanced interactions. Each character exerts a pressure on all others, there are cyclic relationships that are hard to reason about, but this method lays it bare for you, showing approximately how the whole thing might pan out.

High-Skill Sackboy Frequencies

Just to get some intuition that we're moving in the right direction, let's look at some actual data on Sackboy play. The average player didn't take him too seriously; playing according to advantage is definitely a high-skill practice. But look what was happening at the top end of players. (Where, admittedly, Sackboy was stronger.) By the end of the first week, players were already playing a ton of Sackboy. Within three weeks, he completely dominated the meta. We patched him of course, but with metagame bounds we could have known by the first day that he needed patching! The matchups didn't change much after the first couple days; it was the meta that was changing; more and more players realizing it was Sackboy or be Sackboy'd. Furthermore, when we prepare a patch, we don't just need to put it out there. We can run playtests of Sackboy vs each opponent, refill that row of the matrix, and get totally new metagame bounds before ever testing it on players.



Just to get some intuition that we're moving in the right direction, let's look at some actual data on Sackboy play. The average player didn't take him too seriously; playing according to advantage is definitely a high-skill practice. But look what was happening at the top end of players. (Where, admittedly, Sackboy was stronger.) By the end of the first week, players were already playing a ton of Sackboy. Within three weeks, he completely dominated the meta. We patched him of course, but with metagame bounds we could have known by the first day that he needed patching! The matchups didn't change much after the first couple days; it was the meta that was changing; more and more players realizing it was Sackboy or be Sackboy'd. Furthermore, when we prepare a patch, we don't just need to put it out there. We can run playtests of Sackboy vs each opponent, refill that row of the matrix, and get totally new metagame bounds before ever testing it on players.



So you can try this method out yourself, with your own matchups, or anybody else's. Download the tool here, at https://github.com/Blinkity/metagame. It's extremely simple and might even save you some disasters. And I think that's the big picture that I want us to remember. We could just keep throwing stuff out there to see what sticks, listening to player's concerns of the moment. But we end playing whack-a-mole with a game's asymmetry, nerfing something just for something else to pop up and dominate. But games are people's livelihoods now, and we owe it to them and ourselves to take them seriously. Let's try to look more deeply at the structure of a game. Matchups charts form a surprisingly robust, rich view of the nature of an asymmetric competitive game. There are a lot of secrets to uncover from them, and metagame bounds are just one method of more I've uncovered and many I haven't. It's ironic that we've never really applied game theory to this one piece of a game that's really simple enough to be understood by it. Of course, humans are weird, we'll never fully understand how they'll play. But if we can at least understand how they should play, then for the best of the players that's going to be a pretty good start. Try it out and let me know what you uncover!

https://github.com/Blinkity/metagame

Thanks!

@blinkity

ajaffe@gmail.com