

Compassionate Design of Network Platforms: Evaluating Proactive Moderation in *Sky*

Jessica “jynnie” Tang

KEYWORDS.

compassion, proactive moderation, platforms, social interactions, MMOGs

INTRODUCTION.

In recent light of platform scandals with moderation and algorithmic oppression, platforms have attempted to better design their systems to build healthy, inclusive communities. The call for compassionate moderation (Gillepsie 2018) asks if platforms can be designed on ideal morals and facilitate compassionate social interactions. *Thatgamecompany* takes on this challenge in its latest game *Sky: Light Awaits* and hopes to design a compassionate social platform. As I began playtesting the game, I was intrigued by the care that went into its design and the weaving of game and platform mechanics. In playing, I wondered as well if designed mechanics could push a platform towards “compassion and generosity” (*thatgamecompany*, About Sky 2018).

In this essay, I’ve chosen to study a massive multiplayer online game (MMOG) and its proactive moderation design to deal with toxicity and harassment. I’ll interrogate what the game’s design goal is by analyzing the promotional material *thatgamecompany* has published as well as give an overview over the game. I’ll then be doing a close reading of the game mechanics and the affordances and incentives they provide in promoting compassion and discouraging harassment. Finally, drawing on

my participatory observation of the community's engagement with those mechanics, I'll be evaluating *Sky*'s "compassionate design".

As of the time of the writing of this essay, only the beta version 0.3.9 of *Sky* has been released, and it is this version of the game that I will refer to and analyze. Readers should note that the game and its mechanics may change between the writing of this essay and the game's final release. I make no claims to how the community will evolve or what the final gameplay will be like. However, analysis of the current version's design sheds a lot of light on the benefits and limitations of this implementation of "proactive moderation" (Lo 2018).

Before I dive into discussing *Sky*, I want to speak to the importance of video game platforms within the larger discussion on network platforms.

WHY PLAYFUL PLATFORMS.

Video games have been connecting players together on MMOGs and networking communities on a large scale as social media platforms do. Despite the fact that many MMOGs are social platforms, discussion of MMOG networks have been largely missing from the discussion of algorithms and social platforms in media studies. Perhaps this stems from beliefs that the fantasy or playfulness of a game somehow renders the social networks in them fundamentally different or incomparable to those on what we may typically refer to as a social network platform (such as Facebook or Twitter). Or perhaps the concept that social connections made in games are not as valuable or as "real" lingers. But regardless of what model one views the networked sociality in video games as, it cannot be disputed that games host participatory, digital publics where players in a community interact. Game platforms face the same design problems of moderation, of mediation of toxicity, of

community building as other social networking platforms do. Our understanding of moderation and community building efforts from developers, designers, and users of platforms can become much richer by seriously considering MMOG platforms, or what I will call broadly refer to as playful platforms.

WHAT KIND OF SOCIAL.

On the official media outlets for *Sky*, the game is referred to as “a social adventure experience” that hopes to “create a new experience that promotes compassion, altruism and teamwork within its audience base” (thatgamecompany, About Sky 2018). The game is characterized as “social” and an “experience” over a “game”, eschewing any competitive implications. Additionally, the game endeavors to evoke more than just collaboration between players. It wants to evoke “compassion” and “altruism”. In another official statement, gameplay is depicted as “hold[ing] hands with your loved ones as you fly together through a desolate sky kingdom, where generosity and compassion is key to lighting your path” (thatgamecompany, Sky: Light Awaits 2018). Here, compassion is not only embedded in play, but also is “key” to progression. This illuminates some sort of designed in-game benefit for interacting compassionately with others.

The idea of compassion comes up again and again in promotional materials for *Sky*, so let’s unpack that concept. As defined by Merriam-Webster, compassion is the “sympathetic consciousness of others’ distress together with a desire to alleviate it”. It reflects an attitude that considers others with care and an impetus to aid. So, *Sky* is asking for something more than an act of collaboration from its players, it’s asking for caring, emotional engagement. And in trying to promote a compassionate attitude, *Sky* must also discourage harassment and toxicity.



Promotional material of lighting candles for Sky

On a very shallow level, *Sky*'s imagery portrays compassion as “spread[ing] light” (thatgamecompany, *Sky: Light Awaits* 2018). This is literally rendered in the main game mechanic of lighting and gifting candles. The image suggests warmth, aid, and caring. But imagery is not truly enough to promote compassion or discourage grieving. And *Sky* employs more than imagery in trying to build its ideal community, it proactively moderates player behavior. Drawing on moderation work done by Claudia Lo and that done by Joseph Seering and peers (Lo 2018) (Seering, Kraut and Dabbish 2017), I am using *proactive moderation* as the preventative and promotive work done to moderate the kinds of behavior in a community. This differs from reactive moderation which takes action on content after-the-fact, such as reporting players. I have expanded the proactive model of moderation from Lo's model to include the moderation work done by platforms (Lo 2018). Under this framework, I'll be analyzing four major design choices of *Sky* and their role in proactive moderation.

CURATED EXPRESSIONS.

One of the basic way to proactively moderate is to restrict general user capabilities or modes of expression. A very base example being removing chat affordances limits the ability of users to harass

through chat. In *Sky*, players ability to express emotions through animations have been curated. Players can emote, which is the performance of an expressive animations by a player’s avatar. Emotes are animations that only involve one player’s avatar and often function like reacts on Facebook. Some emotes in *Sky* include dancing, pointing, and angrily stomping. Another animation expression is peer-to-peer interactions. These are direct interactions between two avatars or players such as holding hands, hugging, or chatting. I’ll discuss more specifically how peer-to-peer interactions are mediated in later sections.



Interface of emote options and my avatar backflipping

These two forms of animated expression exist ephemerally in a specific time and location, such that a player who is not with you in the same game location at the same time cannot see you emote. These ephemeral animated actions are curated to moderate player expression. For example, there are no sexual or explicit emotes. In addition, no peer-to-peer action is ostensibly negative. For example, a player can’t slap another player’s avatar. While there are emotes for “negative emotions”, such as anger or sadness, they do not engage with another player. This makes it difficult to direct negative expressions at another player. The ephemerality of these actions also avoids the persistence of negativity.

Curation is not only a process of limitation, but also a process of endorsement. Much like how the review of applications on the Apple Store, a form of proactive moderation, purported authorship and an assertion of values (Gillepsie 2018), curation of content to conform to a certain aesthetic and standard of compassion endorses those processes. And as with the Apple Store, *Sky* presents a set of values. These expressions are, however, co-authored by the platform and the player. The platform has authored player's possible expressions, but players utilizing them participate in their enactment. Under the framework of *procedural rhetoric* (Bogost 2008) (Sicart 2011), a framework that purports participation in a process has the potential to persuade, players may be persuaded to interact compassionately.

CONSENSUAL INTERACTIONS.

Sky also proactively moderates in how it mediates peer-to-peer interactions. I refer to these interactions as peer-to-peer as participants exercise equal power in facilitating the interaction. All players involved in a peer-to-peer interaction are required to agree or consent. For example, I can offer a hug to another player and the other player must also accept the hug before our avatars would participate in the interaction. This consent model situates players on equal footing, deterring player ability to take advantage of another. In addition, the consent model is applied to many mechanics. For example, the mechanic of handholding allows avatars to hold hands and give up control of one's avatar to another player. The mechanic of holding hands is both powerful in its intimate imagery, but also in its act of relinquishing control over one's avatar. It allows players to help one another through levels, but also could lead to avatar hijacking. Which is where the consent model of engagement becomes highly important. The player *must* offer to take control for the other player, and the other player *must* explicitly consent. Furthermore, both players have the ability to disengage at any time.

All direct interactions between players is peer-to-peer and applies the consent model. In order to become friends with another player, both players must consent. Players can be unfriended at any time. The two forms of text communication also apply the consent model. Peer-to-peer chat requires either 1) both players to be friends and accept the unlocking of that affordance between them or 2) all players to sit on a bench and light a candle. Speech is censored (transformed into “****”) to players who have not consented through one of the processes.

The second form of text communication is message candles. Message candles are similar to Post-its in that players can leave a message in a space of their choice and another player can read it. In order to read a message candle, a player must light the candle and then choose to read the message. Message candles are rendered aesthetically differently than other candles so that players cannot be tricked into lighting one. And even if the candle was accidentally lit, the two-step process of consenting to read a message candle ensures messages do not spring up unwarranted.

Applying the consent model forces players to recognize other players as independent, equal actors. This intervention by the platform of establishing players with equal power helps discourage cyberbullying, as cyberbullying thrives off “a perceived power imbalance” (Young, Miles and Alhabash 2018). Additionally, drawing on actor network theory, the platform has been codified with certain standards of morality and ideas of equality that are enforced (Latour 2008).

PSUEDOANONYMITY.

Skay additionally renders player identities psuedoanonymous. Players set names for their friends, resulting in a lack of ubiquitous identity. The player has control over their avatar’s customization,

but its greatly limited to only changing the hairstyle (not color), pants, cape color, and accessories. These customizations are not only curated, but also have to be discovered in the world, and purchased. This limited ability to recognize a player in a universal manner makes it difficult to target another player. Additionally, message candles are authorless unless they were placed by one of your friends, in which case, it displays the name the player has inputted for their friend.

Pseudoanonymity makes it difficult to repeatedly target a player or organize a group of players to harass one player. Targeting a player through speech is difficult altogether. Additionally, the limitations of customizability and identity building makes it difficult to bring offline identities into the game. This limits the ability of other players to recognize the gender, ethnicity, and other aspects of a player's offline body and may proactively moderate harassment that targets the offline body. Kishona Gray and others have done a lot of work on the harassment that occurs with the identification of an offline body, especially marginalized bodies (Gray 2016). In her work, Gray discusses how offline identities are often extrapolated (rightly or wrongly) from voice. As identity building is greatly hindered in *Skj*, it is difficult to weaponize racist and sexist beliefs towards specific players on the grounds of their identity. But it does not deter those views from being expressed.

This fluid, pseudoanonymous identity may have been implemented to help combat toxic attitudes towards offline bodies as well as dampen the ability to target another individual. However, this system has major affordances that allow questionable and toxic behavior to continue. For example, players encountered one message candle that broadcasted "send nudes". I couldn't assess whether the message was harassment or not for not only did it target no one, there was no author to send nudes to. But while it was broadcasted to the void, the sexual message was embedded into the

environment. And there was no way for other players to trace the author of the message and *Skry* has no system for reporting.

The lack of identity also makes it difficult for players to recognize role models or toxic players. The community has no way of identifying or sharing knowledge of other players. Thus, if harassment occurs, knowledge of such harassment cannot be shared within the game's system or reported in-game or even on other platforms.

Lastly, while I do not want to linger too long on this idea as it is not directly relevant to moderation of toxicity in game, the concealment or erasure of offline bodies may be problematic. As Dovey and Kennedy argue, our offline bodies do matter to the cybernetic circuitry that comprises gameplay (Kennedy and Dovey 2006). And as Kishona Gray asserts, recognition of marginalized bodies in digital spaces is a manner of progressing towards breaking the assumption that white, male bodies mostly populate digital spaces (Gray 2016). Additionally, players have power over the names of their friends, exerting a power over others' identities without needing the other player's consent or knowledge of how they've been characterized. By isolating players in their own bubbles, ones that look much like filter bubbles as player's have the power to curate their view of the game, problematic ideas and toxicity can circulate on an individual level.

ECONOMIC INCENTIVES TO BE A GOOD PERSON.

The last design choice of *Skry* I want to touch on is the transactional usage of candles. I have mentioned above that some customizations or interactions need to be purchased or consented upon. These transactions of consent and purchase of customizations all occur with the offering of candles. Generally, in game there exists world candles, which matter to one's progress through the game and

are designed into the world, and trading candles, which solely matter in facilitating social transactions. For example, when friending someone, you must offer them a trading candle. If they accept the candle, you lose one candle and they gain one heart and the friend connection has been made. The hearts one accumulates when another player gifts them a candle are traded for the customizations mentioned in the previous section. Thus, friendship becomes a manner of generating currency to unlock customizations and tools. This may proactively incentivize players to be compassionate and kind to one another so that the other player may gift them a candle. But it also spawned “beggars” who pestered other players to gift a candle – instrumentalizing care.

The entanglement of trading candles in consent and customization purchases makes the proactive moderation system much more complicated. Player motivations of making friends and purchasing customizations for one’s self mix. This becomes further complicated by micro-transactions. While the beta version does not implement micro-transactions, the game has suggested that trading candles will be purchasable through micro-transactions. Members of the beta testing community have expressed concern with this model. One player pointed out the consent model can break down if a player paid for a candle, used that to unlock chat with another player, and the other player turns around and uses the chat to harass the original player. This possibly results in the problematic reading that a player paid for harassment. In this case, the proactive moderation not only did not incentivize compassion, it resulted in a monetary loss and harassment. Additionally, another player noted that such a mechanic plays into narratives of buying friends – a less than healthy idea of friendship.

AN INCOMPLETE MODERATION SYSTEM.

Sky's design implements many proactive moderation measures that enforce healthy interactions between players, such as by enforcing consent. However, perhaps in trying to promote compassion from players on the individual level, the design examined in this essay has drained much of the community's power in moderating. *Sky* heavily relies on proactive moderation to discourage player harassment and toxicity, but has limited the community's ability to reactively moderate and assert its values. The community's ability to retain collective knowledge has been greatly stunted by pseudoanonymity and the lack of a reporting affordance. In response, the community organized itself around Facebook groups, Reddit, and Discord servers. This has allowed users to voice their opinions on acceptable play and social norms. The concerns about micro-transactions and candle begging were shared through these other platforms. A current problem the community is struggling with is message candles. Many players have voiced that message candles clog the world with unnecessary and often trollish messages. Additionally, beyond just trollish messages, the trollish placement of some candles in crucial places has led to disgruntle in the community. However, there's no way in game to reach to the rest of the community, not even through reporting or moving "bad" message candles, and negotiate these values. Instead, the community holds discussions and polls on Facebook imploring both players and the developers to remove message candles.

The community has worked to negotiate around problematic constructions of friendship in the game's design. For example, in recognition of the interpretation of friendships as economic transactions, players acknowledged that fruitful friendships and compassionate play could occur without engaging in the candle system. One player¹ says, "it's interesting that people are thinking of candles as required to interact. In Journey [another game by *thatgamecompany*] we had a wonderful time with companions, with no friendship features as all". While the game has made it difficult for

¹ The player granted me permission to quote them anonymously.

players to intervene in its design, emergent culture organized outside of the game continues to allow players to negotiate meaning and community health in a co-creative process.

I believe *Skyy*'s attempt to design proactive moderation serves as a great example of the possibilities and considerations platforms, in MMOGs or not, can consider. *Skyy* demonstrates systems that utilize curation and procedural rhetoric with an acute recognition of encoded morality. This care in design would be valuable in other platforms as well. However, *Skyy* is perhaps also too technologically deterministic in its hope that no reactive moderation process is needed. Players and culture emergently bend and play with the structures put in place for better or for worse. Thus, *Skyy* also demonstrates the need for users to participate in the community moderation process. Platforms should both consider how they can design systems that both promote compassion and allow users to participate co-creatively.

BIBLIOGRAPHY.

- Bogost, Ian. 2008. "The Rhetoric of Video Games." *The Ecology of Games: Connecting Youth, Games, and Learning* (MIT Press) 117-139.
- Gillepsie, Tarleton. 2018. *Custodians of the Internet*. New Haven: Yale University Press.
- Gray, Kishonna. 2016. "'They're too urban': Black gamers streaming on Twitch." In *Digital Sociologies*, by Jessie Daniels, Karen Gregory and Tressie McMillan Cotton, 355-368. Policy Press.
- Kennedy, Helen, and Jon Dovey. 2006. "Bodies and Machines: Cyborg Subjectivity and Gameplay." In *Game Cultures*, by Helen Kennedy and Jon Dovey, 104-122. Open University Press.
- Latour, Bruno. 2008. "Where are the Missing Masses? The Sociology of a Few Mundane Artifacts." *Technology and Society, Building Our Sociotechnical Future* (MIT Press) 151-180.
- Lo, Claudia. 2018. *When All You Have is a Banhammer*. Cambridge, Massachusetts: MIT.
- Postigo, Hector. 2014. "The socio-technical architecture of digital labor: Converting play into YouTube money." *new media & society* 1-18.
- Seering, Joseph, Robert E. Kraut, and Laura Dabbish. 2017. "Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting." *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*.
- Sicart, Miguel. 2011. "Against Procedurality." *The International Journal of Computer Game Research* 11 (3).
- thatgamecompany. 2018. "About Sky." *Facebook*. Accessed May 18, 2018.
<https://www.facebook.com/thatgamecompany/posts/10155515472541317>.
- . 2018. "Sky Test Gold." May 17.
- . 2018. *Sky: Light Awaits*. Accessed May 18, 2018. <http://thatgamecompany.com/sky/>.

Young, Rachel, Stephanie Miles, and Saleem Alhabash. 2018. "Attacks by Anons: A Content Analysis of Aggressive Posts, Victim Responses, and Bystander Interventions on a Social Media Site." *Social Media + Society* 1-14.