# Groups/Community moderation and participation

## Problem

There are a lot of advantages to the anonymous participation in communities, and our users are really excited about it. Users can express yourself freely without fear of being judged, and without these opinions having impact on their employment and/or relationships. The positive effects of anonymous participation can already be seen on Reddit, even though on Reddit users usually have a single identity and global platform-wide reputation (aka karma) - the engagement and polarity of views expressed on Reddit is much higher than in platforms with a single real-life identity. Participating in different communities under different identities gives people additional choices, of not being associated with some discussions, and yet participating in them. Quoting Oscar Wilde: "Give a man a mask and he’ll tell you the truth."

But anonymous participation comes with a large downside: people are able to abuse anonymity, by posting inappropriate content and comments, spamming group, and making community inappropriate to all other members. That is what happened with Usenet - its growth killed it. Currently we have basic moderation capabilities allowing admins to remove messages of the members and prevent them from sending messages, but for groups with the public link nothing stops them from joining again with a different identity. The only current solution is to make all new members joining without posting permissions, and then manually grant them to the trusted users, but it doesn't scale well to large groups.

## Solution

There are two possible solutions:

- rely on members reputation score, that will be calculated based on their actions and community reactions. Similar approaches are used in Reddit and in Habr.com (a Russian IT community). When a member joins a community, they are assigned 0 or some low score that either prevents posting completely or allows very limited posting. As the member participation is seen by the community as positive (e.g., more positive than negative reactions), the reputation score would grow allowing more participation (longer posts, image posts, more frequent posts, etc.). Community rules can allow transferring some of the reputation score to the new members, so members joining by recommendation can have higher initial score (Reddit doesn't allow it, Habr.com does), but if group admins/owners block a member such reputation transfers should also be blocked.

- for public groups that are registered with the discovery / search servers, these search servers can employ these policies as a condition:
    - discovery server must be granted administrative rights to be able to moderate messages and block members.
    - an advertised link to join the group should be controlled by discovery server (so that the group owners can't make discovery server see a content that is different from the actual group content).
    - group policy within the acceptable range should be required (e.g. reputation scoring, limited participation for the new members, depending on group size, etc.).
    - discovery server would run language and image recognition models to detect inappropriate content and remove it automatically - what is appropriate would be determined by the group policy.

These approaches are not mutually exclusive, most likely they need to be combined.

As the group design remains decentralized, some members can obviously modify the code of the client software to post content in violation of their current permissions. But as the same restrictions would be applied by other clients, they simply won't see this content, and the scoring mechanics can further penalize it.

Both solutions rely on group state being consistent, so some of the improvements for group consistency from the previous RFC are required for them to work reliably.

The actual scoring mechanics require formalization. The initial point can be made the same as on some combination of Reddit and Habr.com ideas, and then tweaked to take into account that the reputation scores are community-wide and group-wide.

Once SimpleX adds identity layer, the discovery server can start supporting the reputation scoring for the identity across multiple communities for people who opt in into it.