This is an interesting paper. It makes a some clear contributions to our growing understanding of how vagueness can emerge and be sustained in signaling games. Unfortunately, I am still not convinced that the authors' equations for the replicator diffusion dynamic correctly reflect their written description of the process and the visualization provided in Figure 1. Below I have attempted to thoroughly explain some reasons for my skepticism. From my point of view, I think the authors have two options. If I'm wrong about the mismatch, then they should include a correct derivation of their dynamic from "first principles," so to speak. This is what I've begun below. Start with payoffs in the game with confused states, and work from the bottom-up to derive the dynamic. If I'm correct about the mismatch, then they should rewrite a few paragraphs early in the paper so that their description of the process matches the mathematics. Even if I'm correct about the mismatch, the authors have still made a solid contribution to the literature and with a non-misleading write-up the results should be published.

Focusing just on sending behavior, the authors give the following equation for their replicator diffusion dynamic on pages 9-10:

$$RDD(\sigma) = D_C(RD(\sigma))$$

where σ is the sender's behavioral strategy, $RD(\sigma)$ is the replicator dynamic on behavioral strategies, and $D_C(\sigma)$ is a function that confuses or scrambles behavioral strategies. According to this equation, each step in the dynamic consists of two parts. First, the population compositions change according to the replicator dynamic. Second, the new compositions mutate or get scrambled according to the function D_C (which for the purposes of this paper is multiplication by a confusion matrix C). In other words, this is the replicator dynamic with confusion/mutation at the level of strategies.

Contrast this with the description of the dynamic provided in Figure 1 and the authors' written description on page 8. According to this account, the replicator diffusion dynamic works as follows. Take a signaling game with a state space T. Nature selects a state t_i at random. The sender then goes to observe the state, but her observation is perturbed by some noise. The element C_{ij} of the confusion matrix gives the probability that true state t_i will be perceived to be t_j. Receivers undergo a similar form of confusion when they go to perform actions. In any case, the vision here is that states and actions get confused, and the population composition changes over time according to the replicator dynamic given the payoffs that result from this confusion/noise.

The problem is that the second description (the account exemplified by Figure 1) does not yield the mathematical equations given on pages 9-10. For evidence of this, we can spell out the written description mathematically from the bottom up. If the sender's perceptions get confused by the authors' C matrix, then we cannot directly use the expected utility functions from page 7, because those reference the true state of the world, which the sender does not have unmediated access to. Instead we need EU functions that take as input the sender's observation. Assuming just two states:

$$\begin{aligned} EU'(m, \text{observed } t_1, \rho) = {} & (\text{probability the real state is } t_1 \text{ given perceived } t_1)\times \\ & (\text{EU of sending } m \text{ given that the real state is } t_1 \text{ and receiver uses } \rho) \\ & + \\ & (\text{probability the real state is } t_2 \text{ given perceived } t_1)\times \\ & (\text{EU of sending } m \text{ given that the real state is } t_2 \text{ and receiver uses } \rho) \end{aligned}$$

The relevant probabilities can be found by Bayes' Theorem. For example, if o_1 stands for sender perceiving/observing t_1, then the first probability is:

$$\begin{aligned} P(t_1|o_1) &= \frac{P(o_1|t_1)P(t_1)}{P(o_1)} \\ &= \frac{C_{11}P(t_1)}{P(o_1)} \end{aligned}$$

And the second probability is:

$$P(t_2|o_1) = \frac{P(o_1|t_2)P(t_2)}{P(o_1)}$$
$$= \frac{C_{21}P(t_2)}{P(o_1)}$$

Now, if we take the behavioral replicator dynamic with these expected payoffs we get the following:

$$RD(\sigma)(m_1|o_1) = \frac{\sigma(m_1|o_1)EU'(m_1, o_1, \rho)}{\sum_m \sigma(m|o_1)EU'(m, o_1, \rho)}$$

By substituting the equation for EU' introduced above (and assuming equiprobable states and two messages) we have:

$$= \frac{\sigma(m_1|o_1)\left[\frac{C_{11}P(t_1)}{P(o_1)}EU(m_1, t_1, \rho) + \frac{C_{21}P(t_2)}{P(o_1)}EU(m_1, t_2, \rho)\right]}{\sigma(m_1|o_1)\left[\frac{C11P(t_1)}{P(o_1)}EU(m_1, t_1, \rho) + \frac{C_{21}P(t_2)}{P(o_1)}EU(m_1, t_2, \rho)\right] + \sigma(m_2|o_1)\left[\frac{C11P(t_1)}{P(o_1)}EU(m_2, t_1, \rho) + \frac{C_{21}P(t_2)}{P(o_1)}EU(m_2, t_2, \rho)\right]}$$

$$= \frac{\sigma(m_1|o_1)\left[C_{11}EU(m_1, t_1, \rho) + C_{21}EU(m_1, t_2, \rho)\right]}{\sigma(m_1|o_1)\left[C_{11}EU(m_1, t_1, \rho) + C_{21}EU(m_1, t_2, \rho)\right] + \sigma(m_2|o_1)\left[C_{11}EU(m_2, t_1, \rho) + C_{21}EU(m_2, t_2, \rho)\right]}$$

Unfortunately, this equation is not the same as the equation the authors present on page 10. Here are two substantial differences. First, the authors take the sum $\sum_j C_{ij}$ which represents observing state j when the actual state is i. This gives the wrong interpretation because we want to know the change of behavior when the sender observes state i (the equation is for $RDD(\sigma)(m_k|t_i)$. In other words, we need C_{21} to show up in the numerator because it is relevant to the payoffs whether or not the sender's observation of state 1 is wrong (i.e., the actual state is state 2). Second, in the author's denominator they have the sum of the average payoffs in observed states t_j. But this sum over Φs (the denominator of the normal replicator dynamic) does not involve the confusion matrix C at all. This is odd because the confusions are very relevant for determining the player's expected payoffs, and $\Phi(t_j, \sigma, \rho)$ tracks the average expected payoff after observing state t_j.