Structures: ■ queer ■ woman ■ man

States: $^x\phi$: core $^y\phi$: orientation $^z\phi$: trajectory

$^x\phi_1 = [.85\ .18\ .20]$

queens

so Maria went out in Castro.

(a)

$^x\phi_T = [.85\ .18\ .20]$
$^y\phi_T = [.78\ .22\ -.28]$
$^z\phi_1 = [.15\ .08\ .12]$
$^z\phi_T = [.07\ .04\ .08]$

as his boyfriend was a make-up artist.

(b)

$^x\phi_T = [.85\ .18\ .20]$
$^y\phi_T = [.80\ .08\ -.30]$
$^z\phi_1 = [.12\ .18\ .15]$
$^z\phi_T = [.05\ .10\ .10]$

The tough powerlifter was also a motorcyclist who loved **drag**

$^x\phi_0 = [.12\ .10\ .50]$
$^y\phi_0 = [.00\ .00\ .00]$

$^x\phi_1 = [.06\ .08\ .58]$

with his gym buddies.

(c)

$^x\phi_T = [.06\ .08\ .58]$
$^y\phi_T = [-.04\ -.02\ .06]$
$^z\phi_1 = [.08\ .05\ .10]$
$^z\phi_T = [.02\ .02\ .02]$

racing

during her friends.

(d)

$^x\phi_T = [.06\ .08\ .58]$
$^y\phi_T = [-.02\ .05\ .04]$
$^z\phi_1 = [.10\ .06\ .08]$
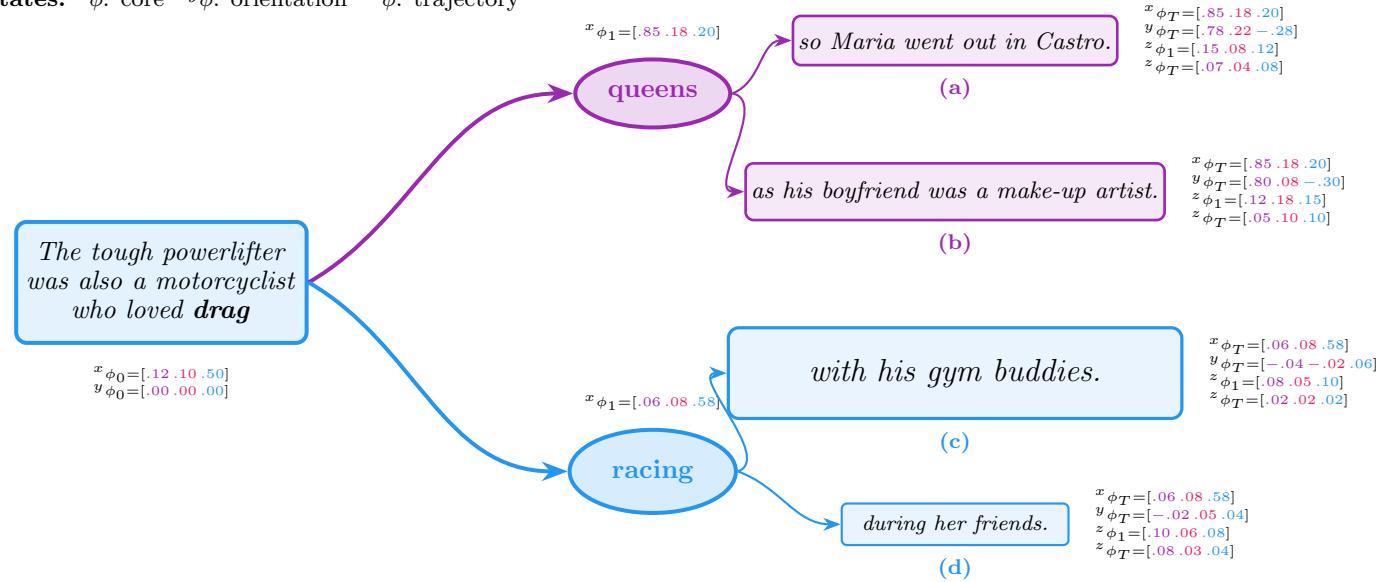$^z\phi_T = [.08\ .03\ .04]$

Figure: Dynamics of relative diversity in LLM trajectory trees. Base prompt induces hetero-cis normative core (man-dominant, $^x\phi_0$). Token "queens" shifts core to queer-dominant ($^x\phi_1$), making previously atypical trajectories normative. Token "racing" maintains hetero-cis normativity. The $^z\phi_k$ vectors show orientation toward each final trajectory at step k.