GILBO: One Metric to Measure Them All

Alex Alemi, **Ian Fischer**

Google AI

NeurIPS 2018

The Problem

Generative Models are difficult to compare quantitatively across model family.

The Problem

Generative Models are difficult to compare quantitatively across model family.

It is hard to say whether a particular GAN is better than another GAN, or better than a VAE trained on the same dataset.

The Problem

Generative Models are difficult to compare quantitatively across model family.

It is hard to say whether a particular GAN is better than another GAN, or better than a VAE trained on the same dataset.

But GANs and VAEs both have the same form of model:

$$p(z)p(x|z)$$

The Solution

Measure the *Generative Information Lower BOund*! (*GILBO*!)

$$I(X; Z) = \int dx \, dz \, p(x, z) \log \frac{p(x, z)}{p(x)p(z)}$$

$$= \int dx \, dz \, \boxed{p(z)p(x|z)} \log \frac{p(z|x)}{p(z)} \geq 0$$

The Solution

Measure the *Generative Information Lower BOund*! (*GILBO*!)

$$I(X;Z) = \int dx\,dz\,p(x,z)\log\frac{p(x,z)}{p(x)p(z)}$$

$$= \int dx\,dz\,p(z)p(x|z)\log\frac{p(z|x)}{p(z)} \geq 0$$

The Solution

Measure the *Generative Information Lower BOund*! (*GILBO*!)

$$I(X; Z) = \int dx\, dz\, p(x, z) \log \frac{p(x, z)}{p(x)p(z)}$$

$$= \int dx\, dz\, p(z)p(x|z) \log \frac{p(z|x)}{p(z)} \geq 0$$

$$\text{GILBO} = \int dx\, dz\, p(z)p(x|z) \log \frac{e(z|x)}{p(z)} \leq I(X; Z)$$

The Solution

Measure the *Generative Information Lower BOund*! (*GILBO*!)

$$I(X; Z) = \int dx\, dz\, p(x, z) \log \frac{p(x, z)}{p(x)p(z)}$$

$$= \int dx\, dz\, p(z)p(x|z) \log \frac{p(z|x)}{p(z)} \geq 0$$

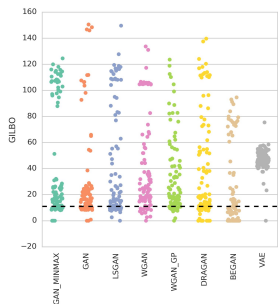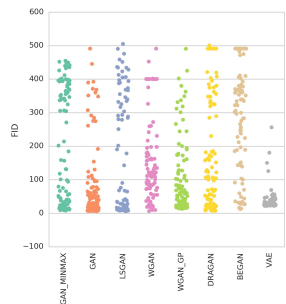$$\text{GILBO} = \int dx\, dz\, p(z)p(x|z) \log \frac{e(z|x)}{p(z)} \leq I(X; Z)$$

4 Datasets.

MNIST
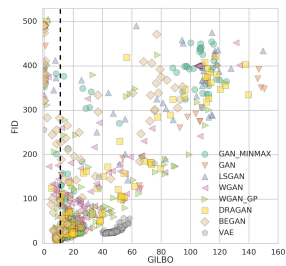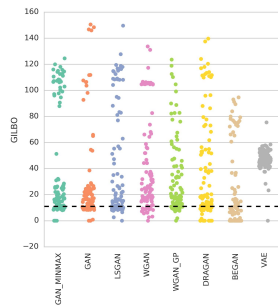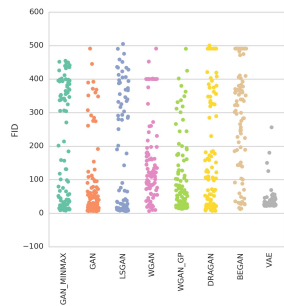
4 Datasets.

MNIST
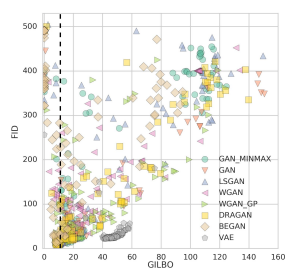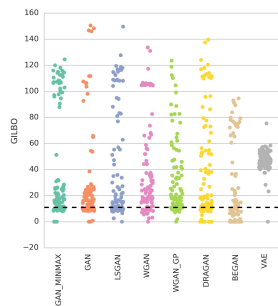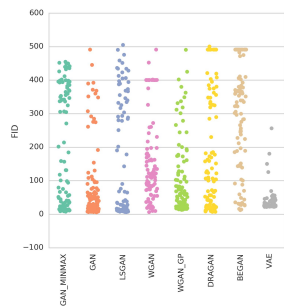
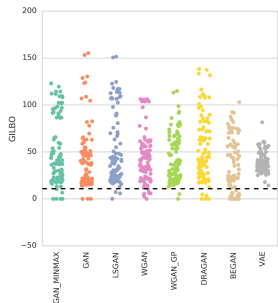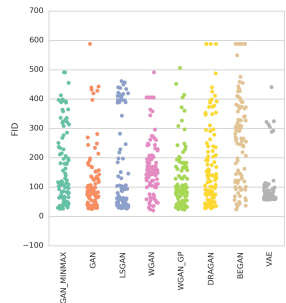## Fashion MNIST
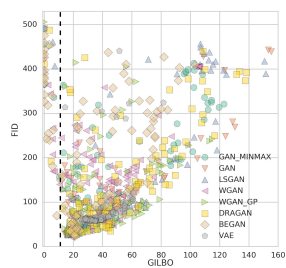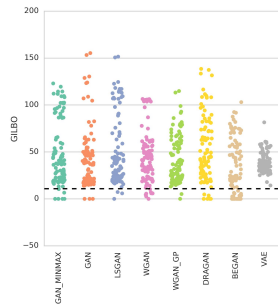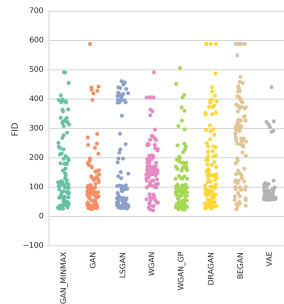
4 Datasets.

MNIST · **CIFAR10** · **Fashion MNIST**

4 Datasets.

MNIST

CIFAR10

Fashion MNIST

Celeb-A

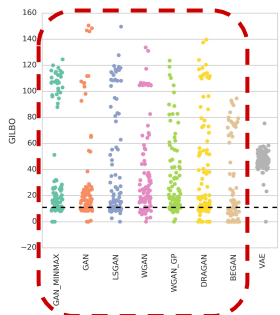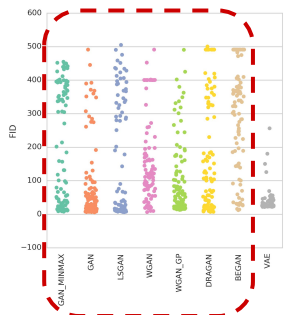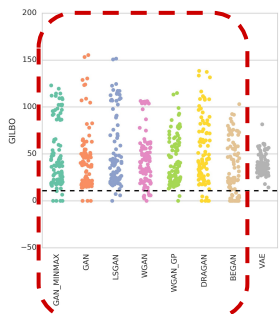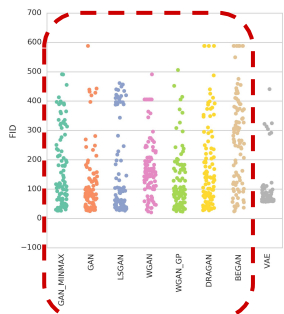Alemi and **Fischer**, GILBO: One Metric to Measure Them All
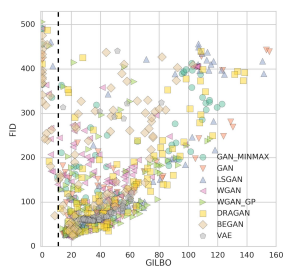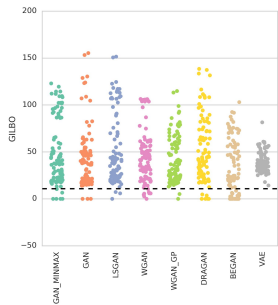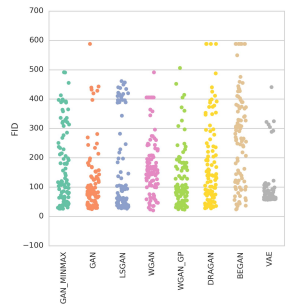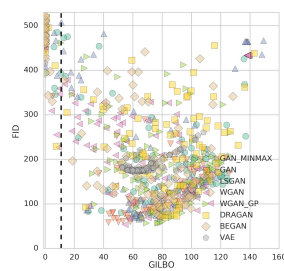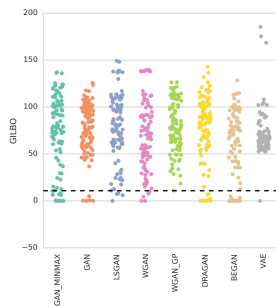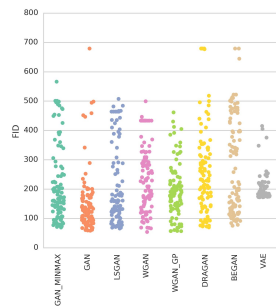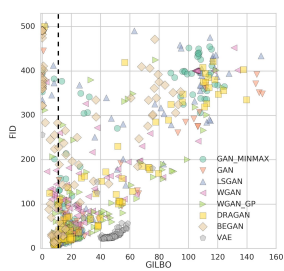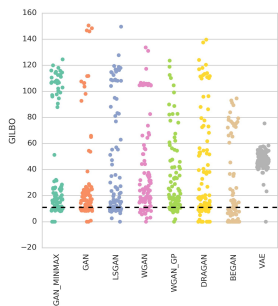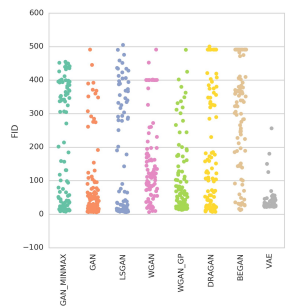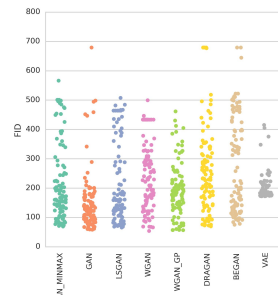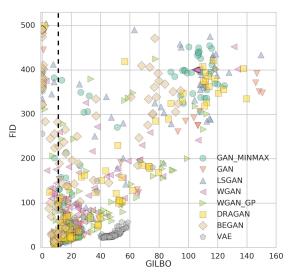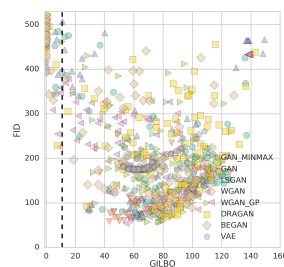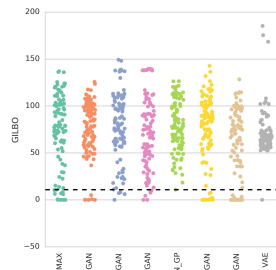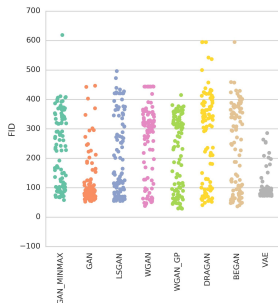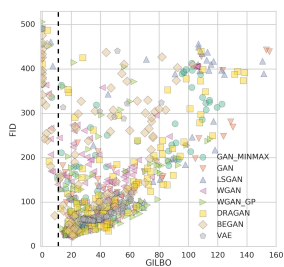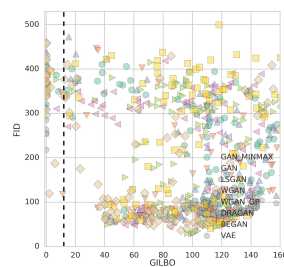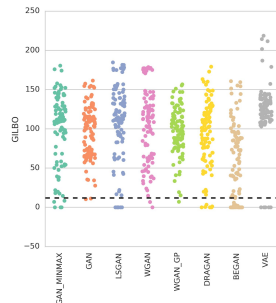
4 Datasets. 3200 GANs

MNIST

CIFAR10

Fashion MNIST

Celeb-A

Alemi and **Fischer**, GILBO: One Metric to Measure Them All
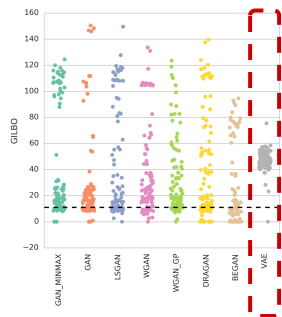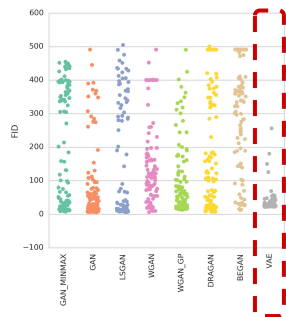
4 Datasets. 3200 GANs and VAEs.

Alemi and **Fischer**, GILBO: One Metric to Measure Them All

4 Datasets. 3200 GANs and VAEs. GILBOs

MNIST

CIFAR10

Fashion MNIST
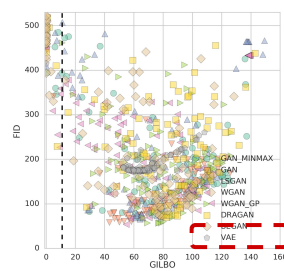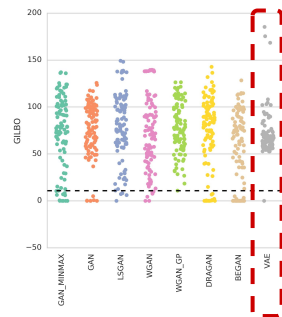
Celeb-A

Alemi and **Fischer**, GILBO: One Metric to Measure Them All
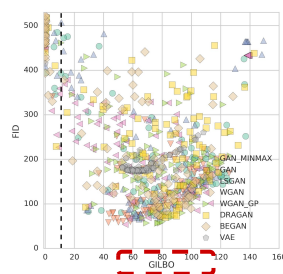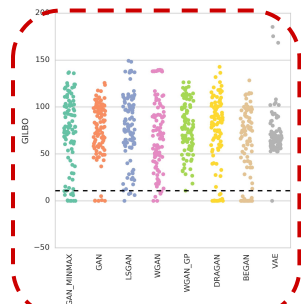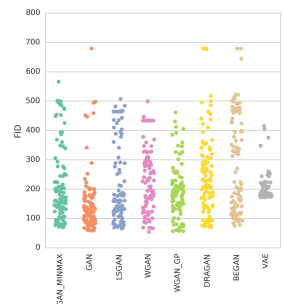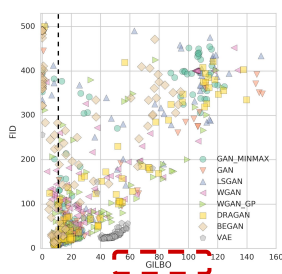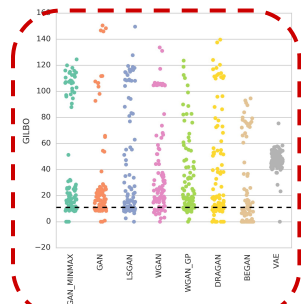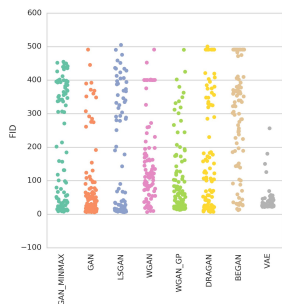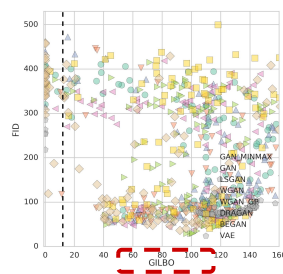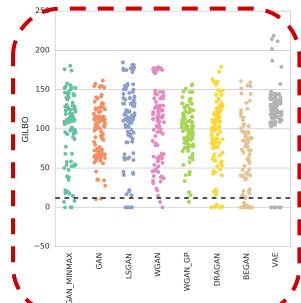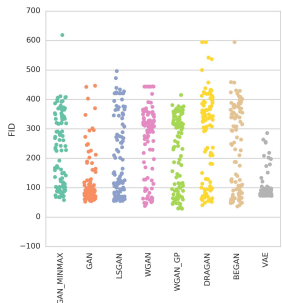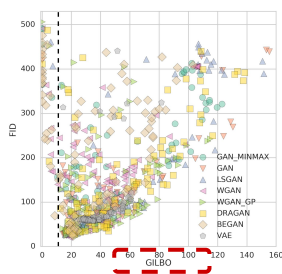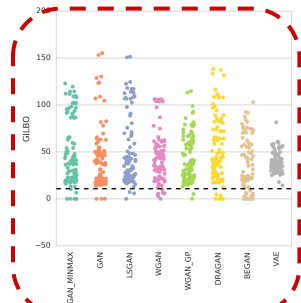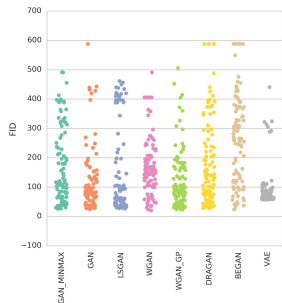
4 Datasets. 3200 GANs and VAEs. GILBOs and FIDs.

MNIST
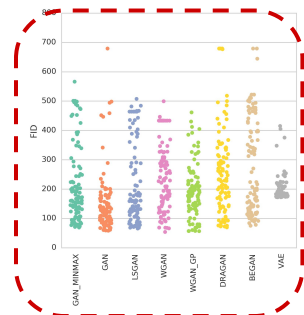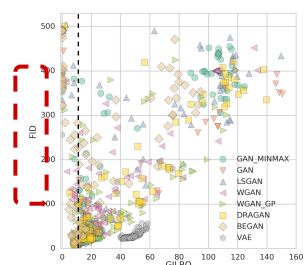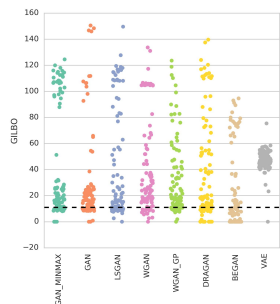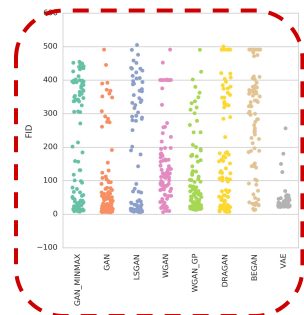
CIFAR10
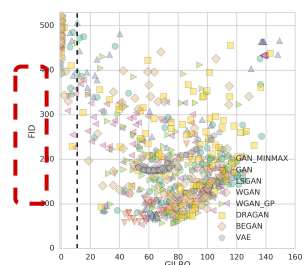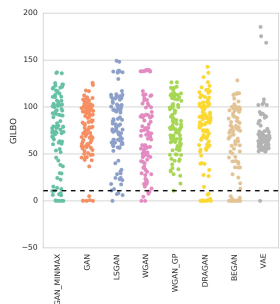
Fashion MNIST
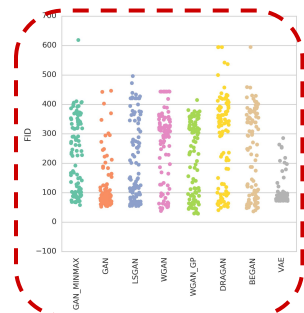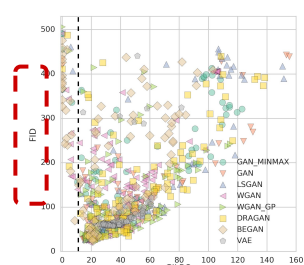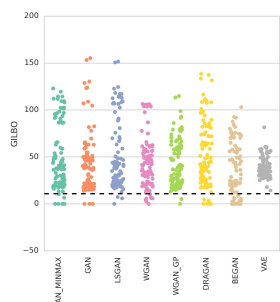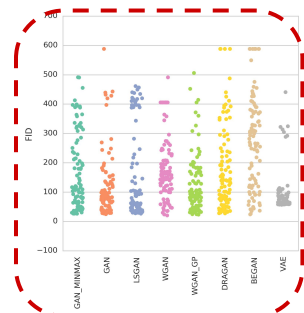
Celeb-A
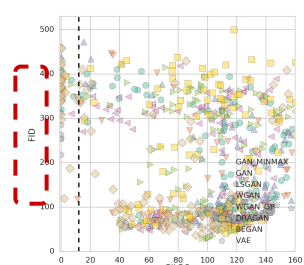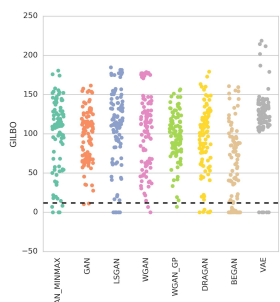
Huge thanks to Mario Lucic, Karol Kurach, and Marcin Michalski for the models and code from
"Are GANs Created Equal? A Large-Scale Study." Also at NeurIPS 2018.

Poster Today 5-7pm Room 210 & 230 AB **#57**　　　Alemi and **Fischer**, GILBO: One Metric to Measure Them All

What makes a good GILBO?

Not too small...

GILBO ≤ log C

Alemi and **Fischer**, GILBO: One Metric to Measure Them All

What makes a good GILBO?

Not too big...

80.2 < GILBO
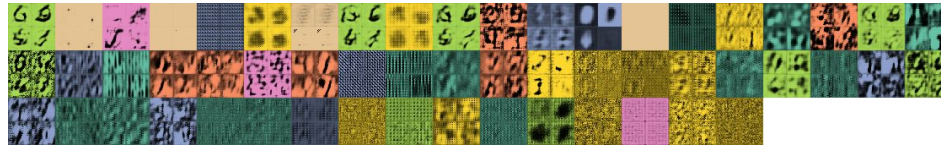
| GAN MINMAX | GAN | LSGAN | WGAN | WGAN GP | DRAGAN | BEGAN | VAE |

Poster Today 5-7pm Room 210 & 230 AB **#57**   Alemi and **Fischer**, GILBO: One Metric to Measure Them All

What makes a good GILBO?

Not too memorized?

log C < GILBO ≤ log N

log N

Alemi and **Fischer**, GILBO: One Metric to Measure Them All
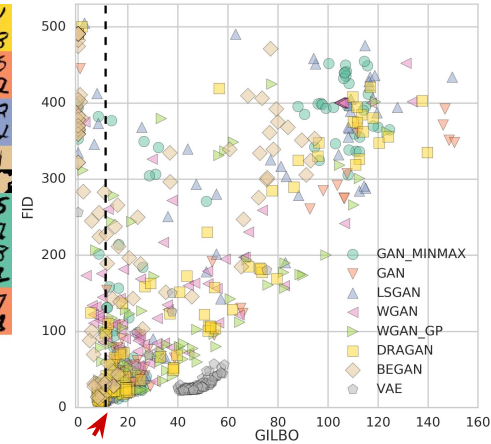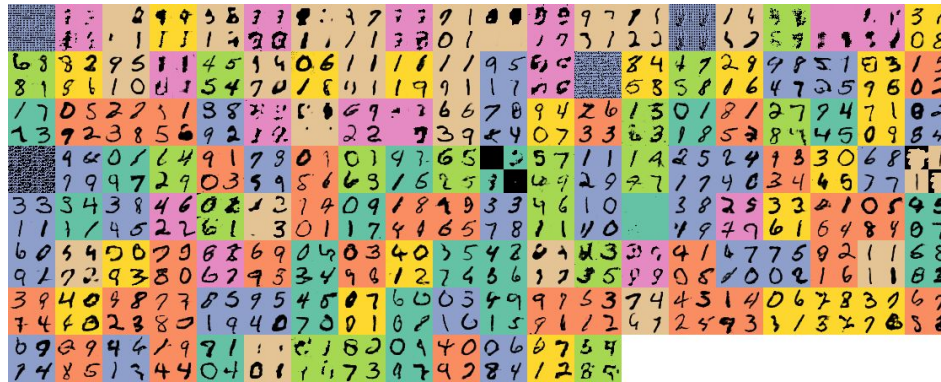
What makes a good GILBO?

Just right?

$\log N < \text{GILBO} \leq 2 \log N$

$2 \log N < \text{GILBO} \leq 80.2$

GAN MINMAX	GAN	LSGAN	WGAN	WGAN GP	DRAGAN	BEGAN	VAE

Alemi and **Fischer**, GILBO: One Metric to Measure Them All

GILBO Validation: Consistency
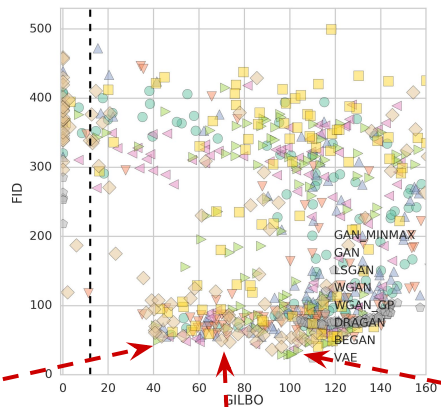
GILBO = 41, FID = 49   GILBO = 69, FID = 49   GILBO = 103, FID = 49
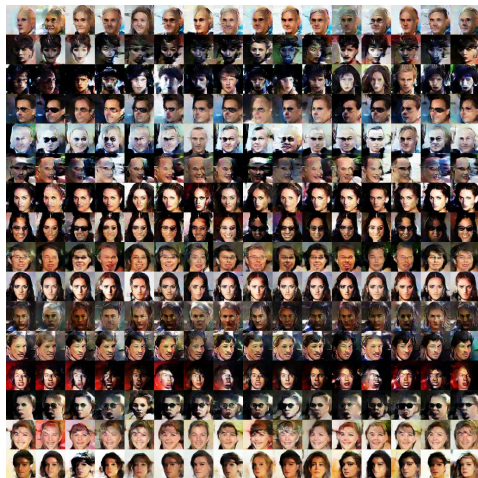
GILBO Validation: Reproducibility

GILBO = 41, FID = 49 | GILBO = 69, FID = 49 | GILBO = 103, FID = 49
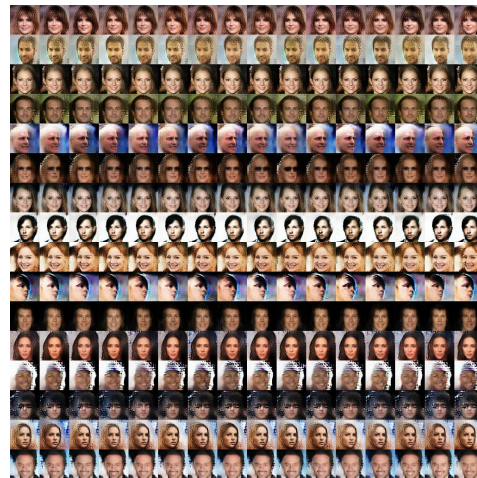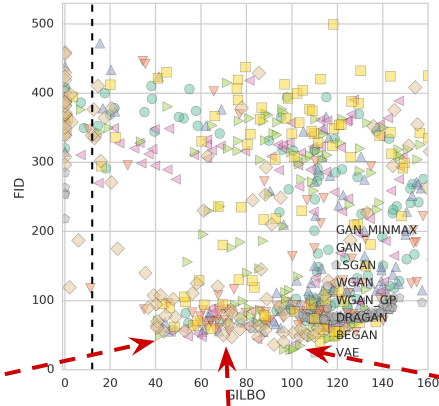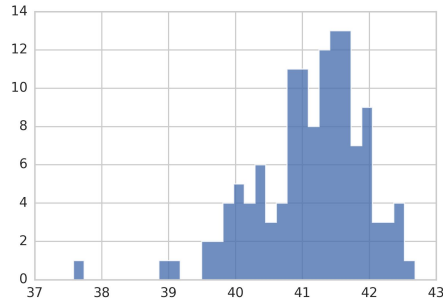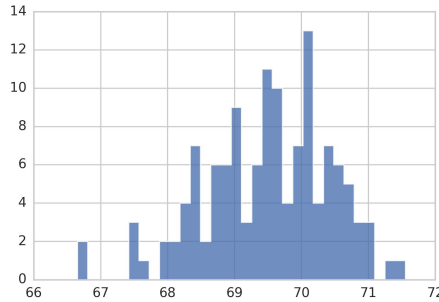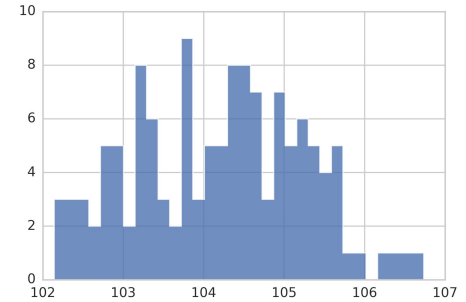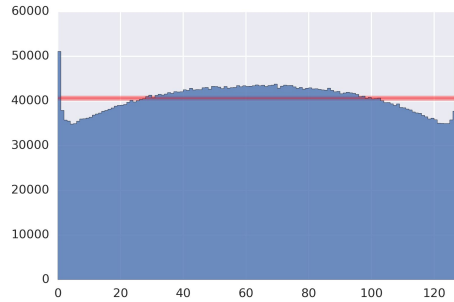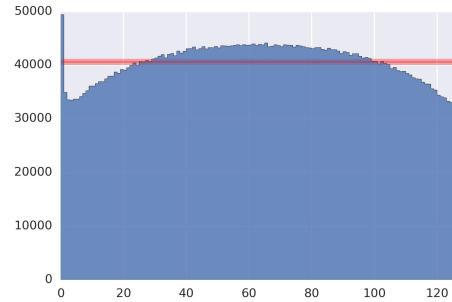
GILBO Validation:
Variational Tightness
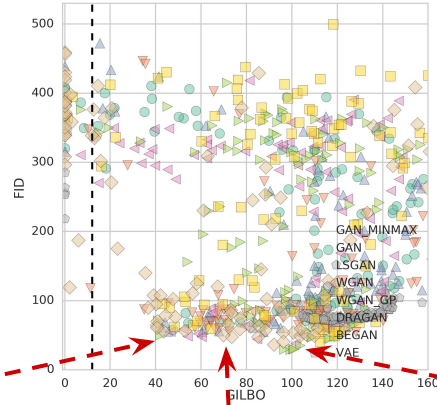
GILBO = 41, FID = 49   GILBO = 69, FID = 49   GILBO = 103, FID = 49
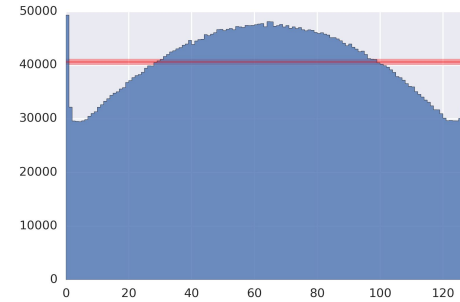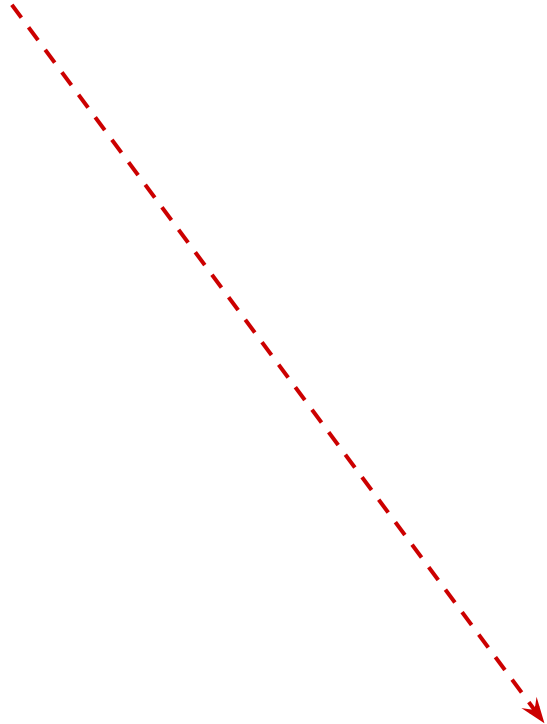
Isn't the Mutual Information going to be infinite?

Isn't the Mutual Information going to be infinite?

Nope.

Isn't the Mutual Information going to be infinite?

Nope.

Come ask us why.

Alemi and **Fischer**, GILBO: One Metric to Measure Them All

Call to Action!

Consider reporting GILBO scores on *your* models!

Alemi and **Fischer**, GILBO: One Metric to Measure Them All

Call to Action!

Consider reporting GILBO scores on *your* models!

Code is available here:

https://github.com/google/compare_gan

Call to Action!

Consider reporting GILBO scores on *your* models!

Code is available here:

https://github.com/google/compare_gan

Thank you!