Quantifying Copyright

James Grimmelmann

March 6, 2017

I. Kolmogorov

Complexity

Five key ideas

- Digital encoding
- Counting bits
- Compression
- Programs
- Conditional complexity

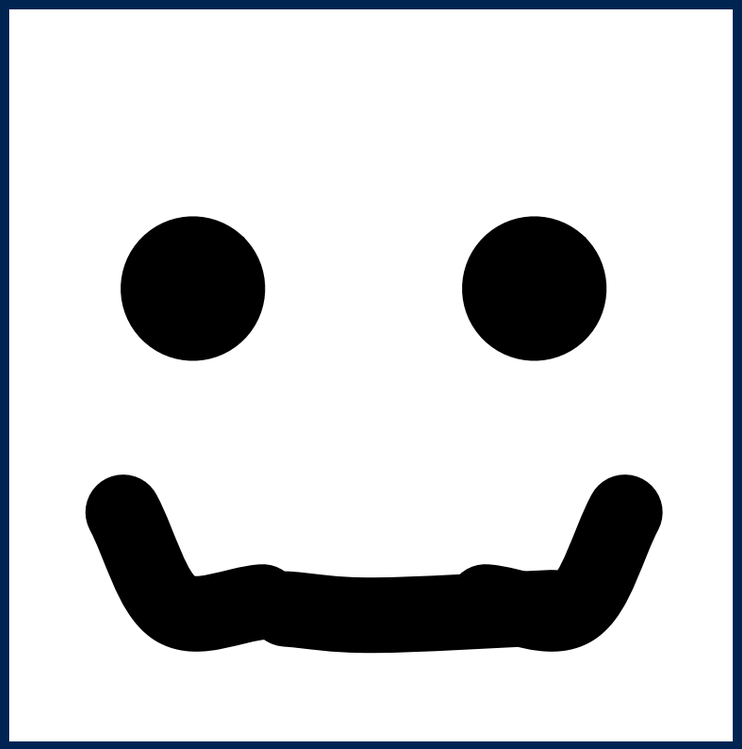
1. Digital encoding

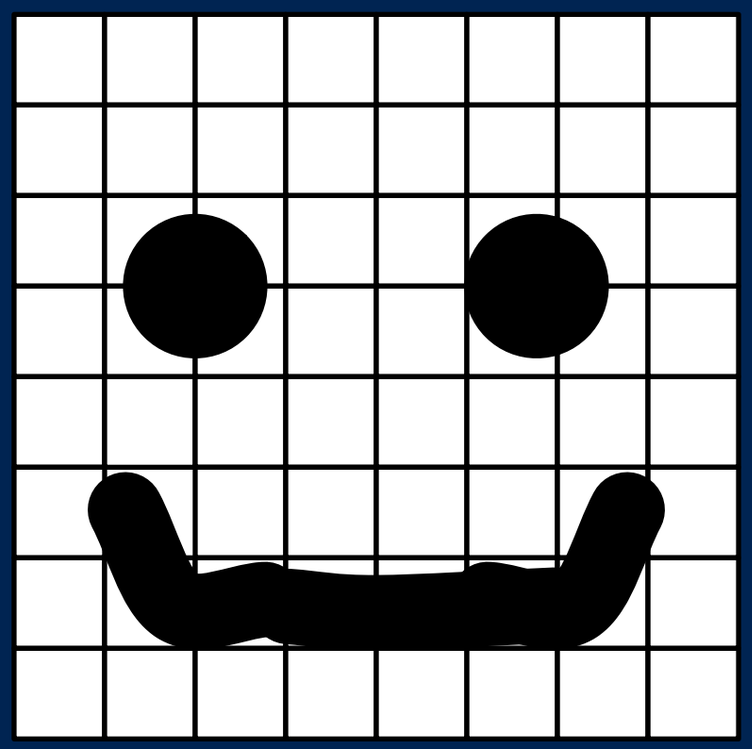
H e l l o !

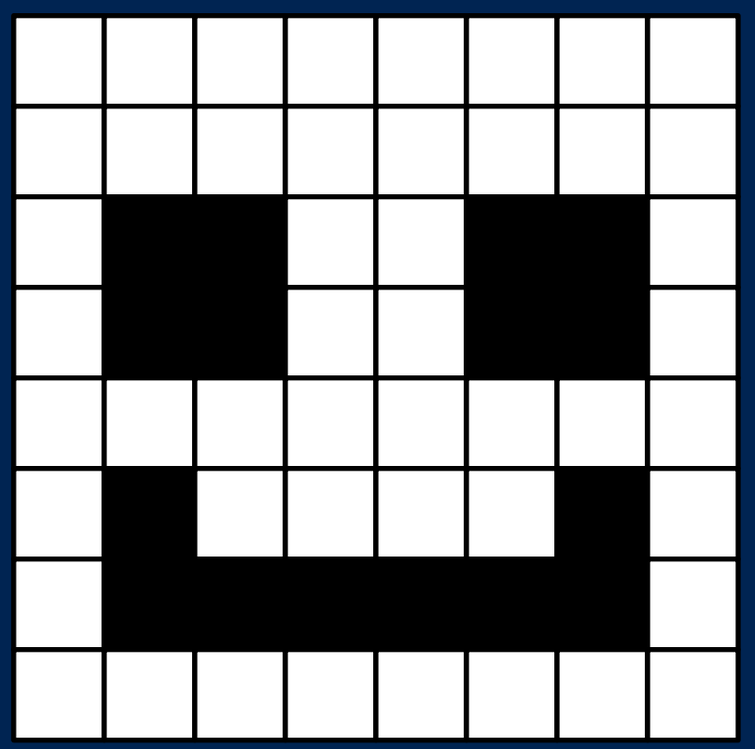
72 101 108 108 111 33

01001000 01100101 01101100 01101100 01101111 00100001

1. Digital encoding





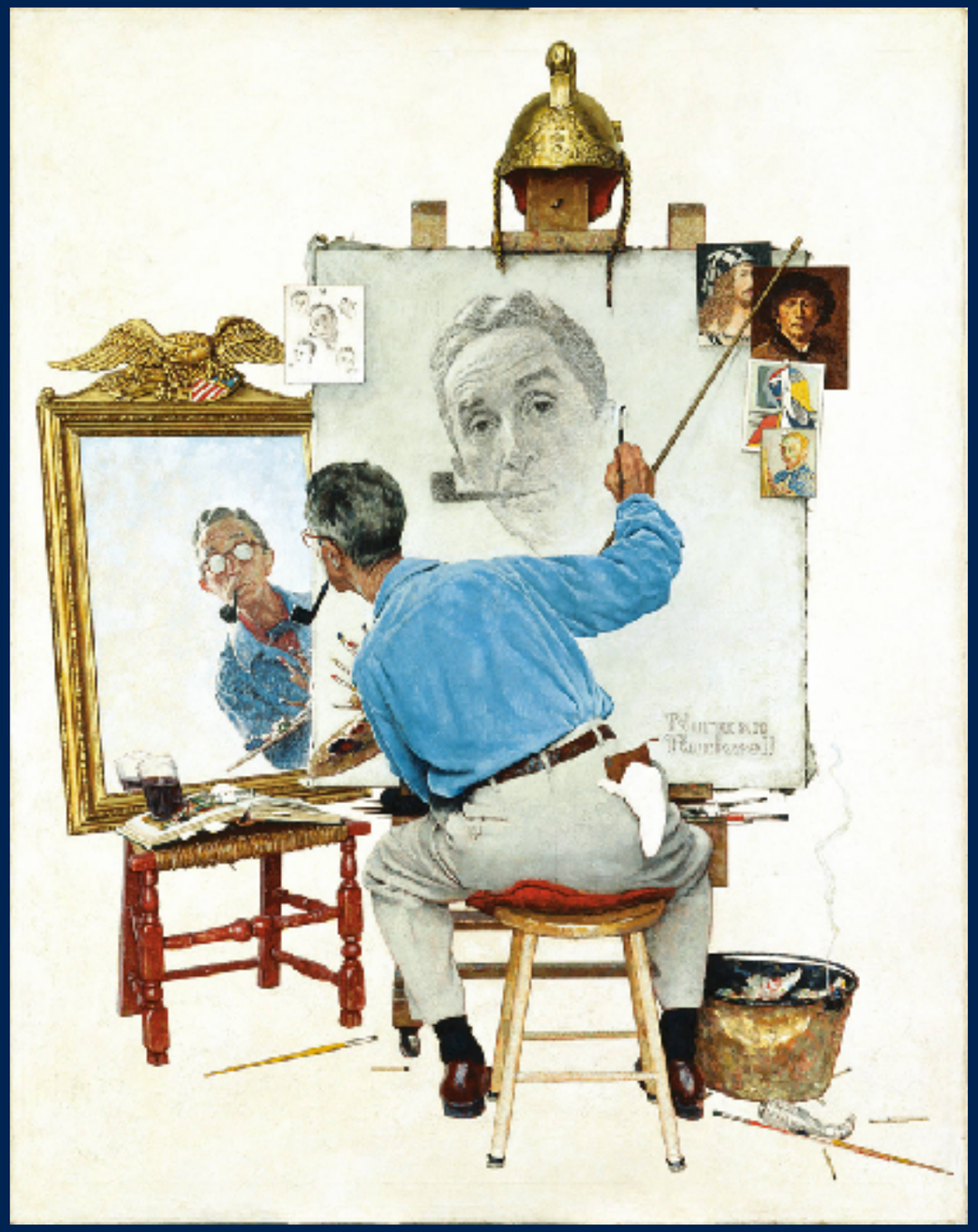


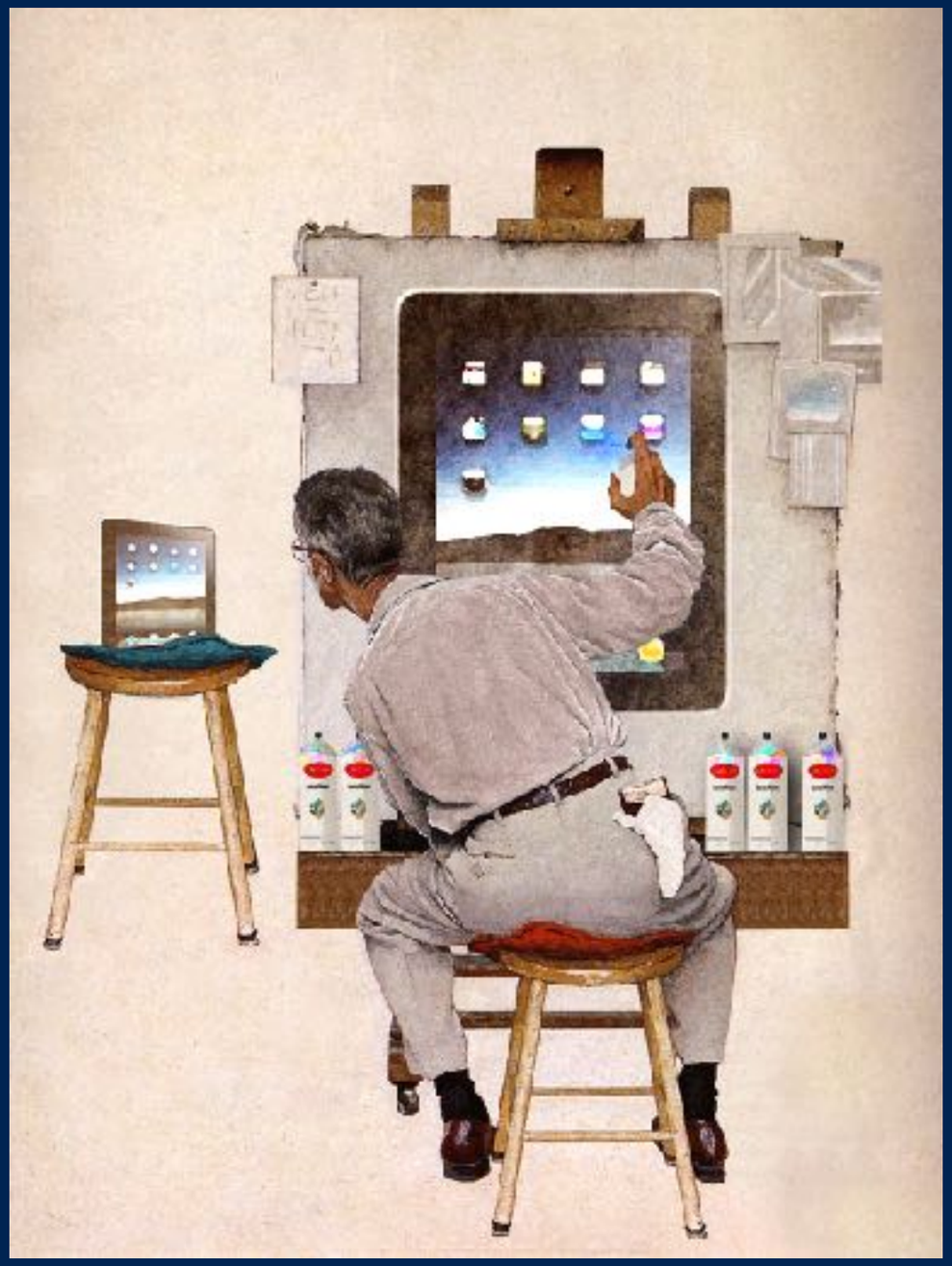
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 0 0 1 1 0 0 1
1 0 0 1 1 0 0 1
1 1 1 1 1 1 1 1
1 0 1 1 1 1 0 1
1 0 0 0 0 0 0 1
1 1 1 1 1 1 1 1

1. Digital encoding




1. Digital encoding





2. Counting bits

You Got the Right One, Uh-Huh

232 bits

2. Counting bits

```
Thou still unravished bride of quietness,
Thou foster-child of silence and slow time,
Sylvan historian, who canst thus express
A flowery tale more sweetly than our rhyme
…
“Beauty is truth, truth beauty,-that is all
Ye know on earth, and all ye need to know”
```

17,544 bits

3. Compression

```
Happy_birthday_to_you←
Happy_birthday_to_you←
Happy_birthday_dear_X←
Happy_birthday_to_you←
```

704 bits

3. Compression

```
1:Happy_birthday
2:*1_to_you←

*2*2*1_dear_X*2
```

368 bits

4. Programs

$$K(w) = \min l(p) : p()=w$$

4. Programs

```
XXXXXXXXXX (10 Xs)                    80 bits

XXXXXXXX…XXXXXX (100 Xs)             800 bits

XXXXXXXX…XXXXXXXX (1000 Xs)         8000 bits

XXXXXXXX…XXXXXXXXXX (10000 Xs)     80000 bits
```

4. Programs

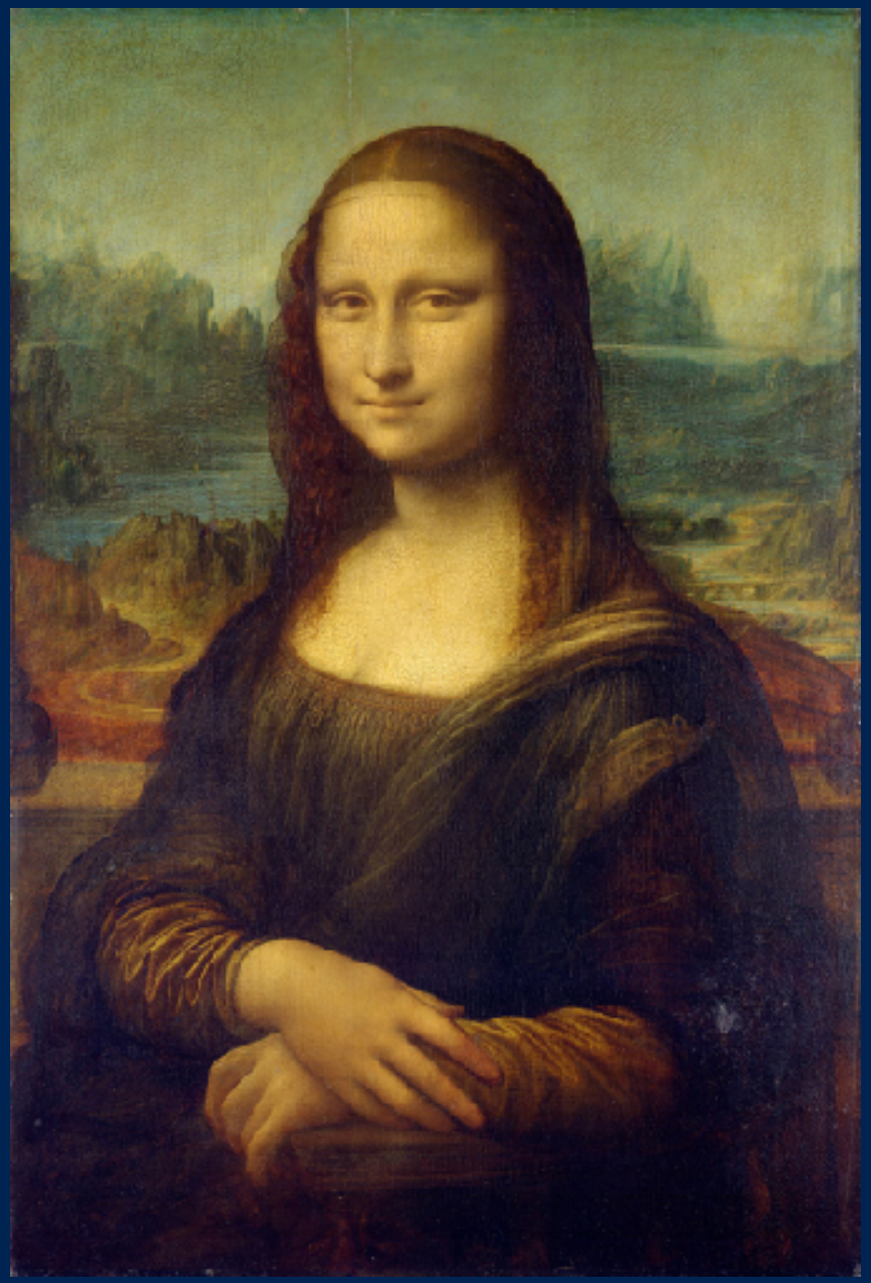
```
XXXXXXXXXX (10 Xs)                    80 bits
for($i=0;$i<10;$i++){print"X"}       240 bits

XXXXXXXX…XXXXXX (100 Xs)             800 bits
for($i=0;$i<100;$i++){print"X"}      248 bits

XXXXXXXX…XXXXXXXX (1000 Xs)         8000 bits
for($i=0;$i<1000;$i++){print"X"}     256 bits

XXXXXXXX…XXXXXXXXXX (10000 Xs)      80000 bits
for($i=0;$i<1000;$i++){print"X"}     264 bits
```

5. Conditional complexity

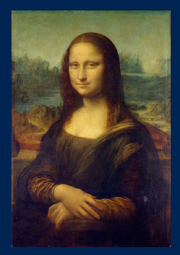


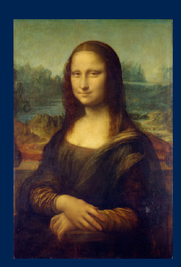
5. Conditional complexity

$$K(x|y) = \min l(p) : p(y)=x$$

5. Conditional complexity

$K($  $) =$ high

$K($  $|$  $) =$ medium

$K($  $) - K($  $|$  $) =$ high

$K($  $|$  $) =$ high

$K($  $) - K($  $|$  $) =$ low

II. Copyright

An unworkable proposal

- Use $K(x)$ to measure expression
- Use $K(x|y)$ to measure similarity
- *Virtue*: common metric among different types of work (poems, movies, songs, etc.)
- *Virtue*: filtration of unoriginal similarity

Objections

- *Objection*: *K* is uncomputable
- *Objection*: *K* ignores psychology and aesthetics

A revised proposal

- *K* cannot show that a work is expressive, but it can show that a work *is not* expressive
- Use *K* as a first step to ask whether a work is complex enough to be potentially copyrightable, or whether two works have enough similarity for infringement
- Second step: fact-finding about ordinary lay audience's reactions
- Maybe the Ninth Circuit has it right!