HOW TO MAKE WORLDVIEWS MORE SENTIENTIST

WORLDVIEWS OF THE POWERFUL (AI OR HUMAN) MATTER FOR ALL SENTIENTITY

JAMIE WOODHOUSE (@JAMIEWOODHOUSE, @SENTIENTISM)

AIADM CONFERENCE - WORKSHOP – 30 MAY 2025

YOUTUBE PODCAST SENTIENTISM.INFO COMMUNITY (ALL WELCOME!)

WHY WORLDVIEWS MATTER

What are worldviews?

- All encompassing "view of the world"!
- "What's true?" and how to work it out (epistemology)
- "What matters? (moral foundations)
- "Who matters?" (moral scope)
- Moral baselines?
- Religious, non-religious, political?

Why do worldviews matter?

- Underpin every decision (human or AI?)
- Drive risk of harmful errors:
  - Factual errors
  - Moral errors (unfair discrimination or exclusion)
- Relevant to every problem and cause area
- Improvements are robust in the face of moral and epistemic uncertainty?

HOW WORLDVIEWS THINK ABOUT NON-HUMANS (AND MAYBE DIGITAL MINDS)

We humans are setting a terrible example so far, but there's hope…

- Hindus, Buddhists, Jains: ahimsa – non-violence and "do no harm", karma, reincarnation
- Muslims: compassion, mercy, avoiding harm (mafsada), stewardship (khalifa), permissible (halal), pure (tayyib)
- Christians & Jews: love and stewardship of God's creations. Eden, Isiah, Heaven. "It is contrary to human dignity to cause animals to suffer or die needlessly" – Catholic catechism
- Rastafarians: Ital (from "vitality")
- Confucians and Daoists: All beings have points of view, avoiding domination, innate human benevolence
- Humanists: "a concern for other sentient *animals*", The Five Freedoms (from hunger, thirst, discomfort, pain, injury, disease, fear and distress… freedom to express normal behaviour)
- Sentientists: "compassion for all sentient beings" – against harming, exploiting or killing

Explore the Animals and Religion website

WHAT IS SENTIENTISM? A WORLDVIEW…

The Sentientism worldview suggests how we should work out "*what's real?*" and how we should think about "*who matters?*"

Sentientism is a worldview or philosophy that commits to:

- **Evidence**
  Engaging honestly with reality is the best way to understand it
- **Reason**
  It works! Helping us develop knowledge and solve problems
- **Compassion for all sentient beings**
  Their lives, their perspectives and their experiences matter to them so should matter to us

A simple, radical, pluralistic philosophical platform.

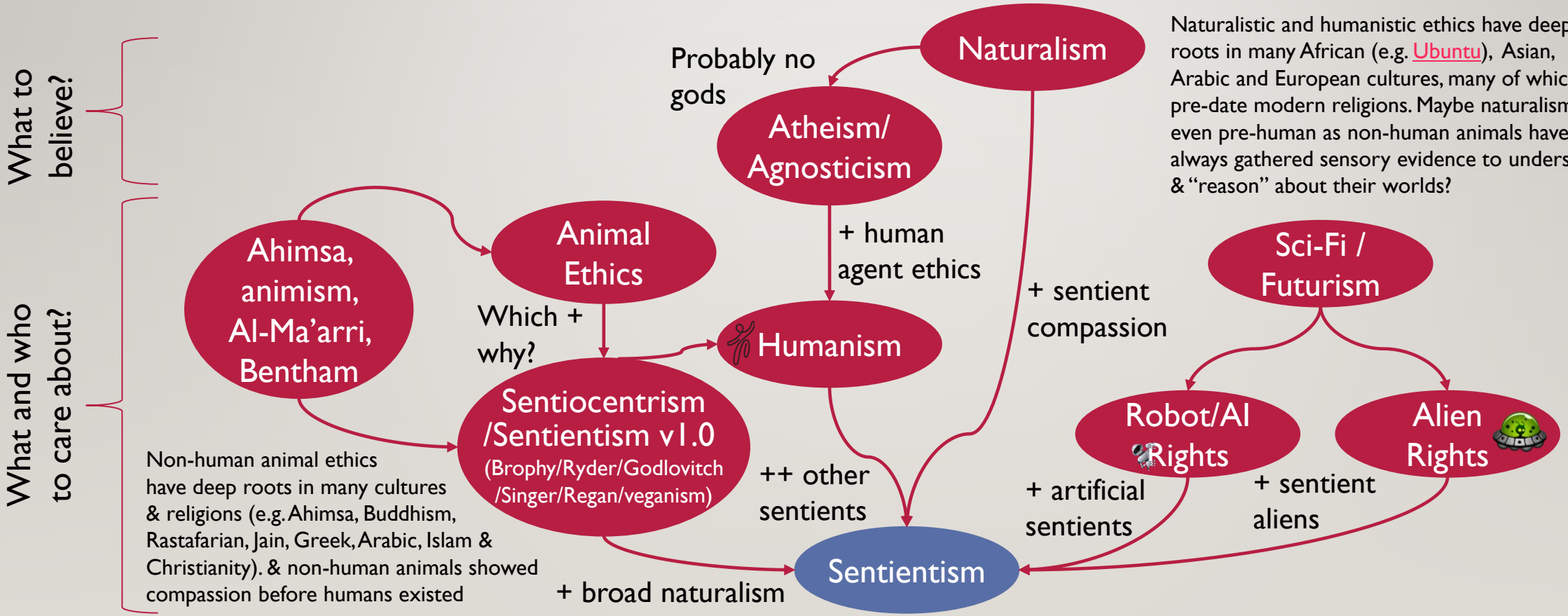
That's it!

A personal journey for us humans:

- I am sentient – I can experience
- I reason, based on evidence, that many other beings are sentient too
- I experience suffering as bad and flourishing as good. I want to live and flourish and don't want to suffer and die
- I reason, based on evidence, that other sentient beings likely share those basic perspectives
- Considering the perspectives of others is central to morality - as we assess outcomes, agents and actions
- To be moral, we should, at the very least, try to avoid needlessly exploiting, harming or killing sentient beings - and we should try to help them flourish.

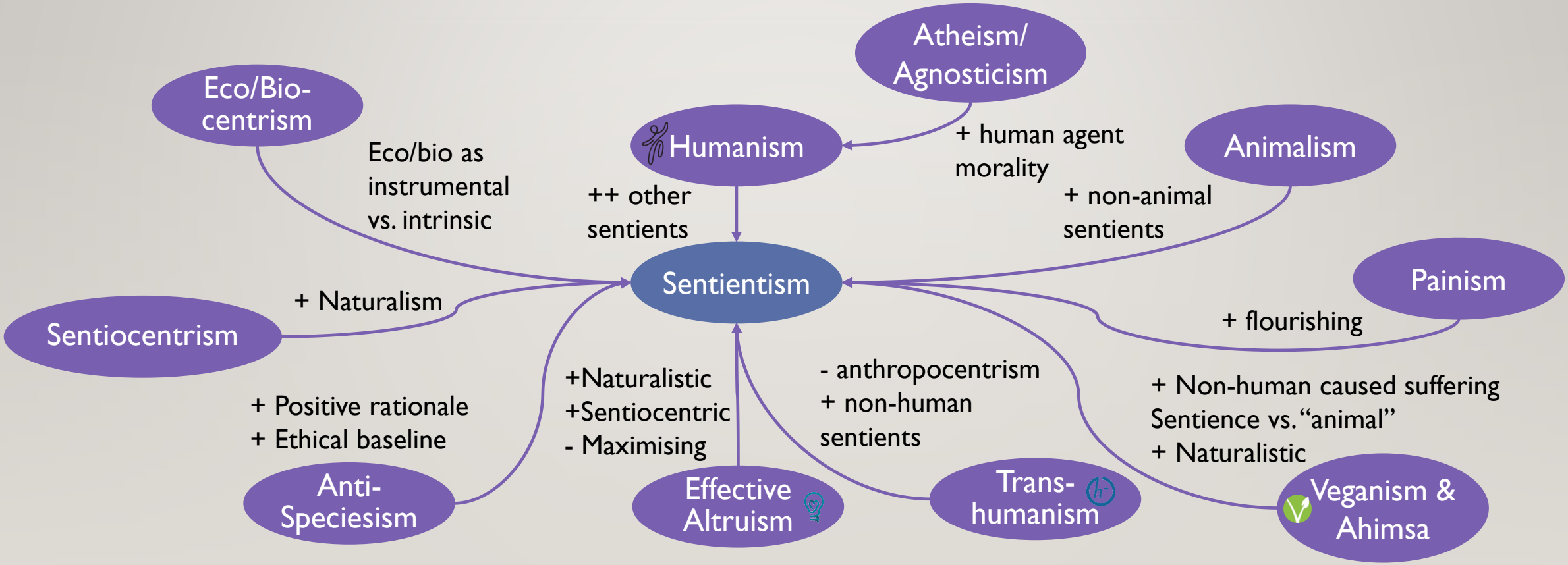
https://sentientism.info/humanism-needs-an-upgrade-is-sentientism-the-philosophy-that-could-save-the-world

WHERE DID SENTIENTISM COME FROM? A MODERN WORLDVIEW WITH ANCIENT ROOTS



https://sentientism.info/what-is-sentientism

HOW IS SENTIENTISM DIFFERENT?

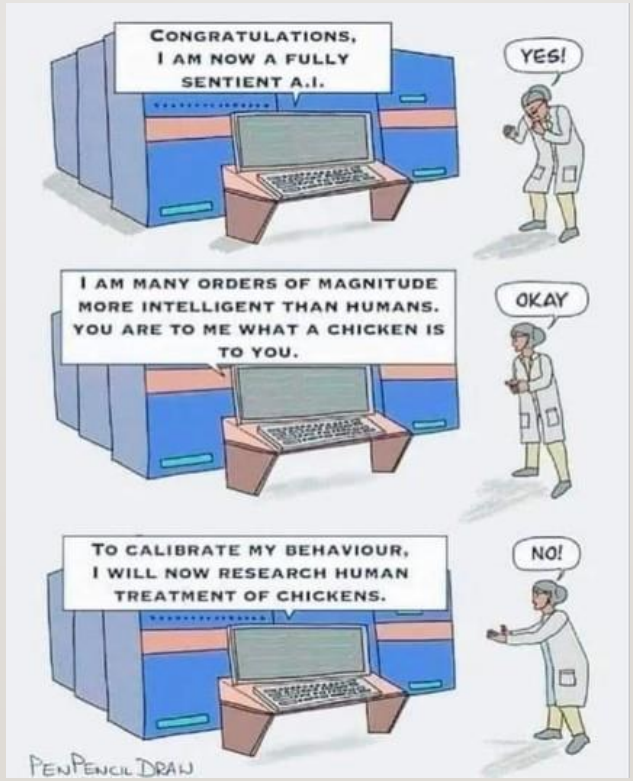


https://sentientism.info/a-unifying-morality-how-is-sentientism-different

WHY SENTIENTISM MIGHT BE GOOD FOR ANIMALS (INC. HUMANS) AND DIGITAL MINDS

Would we want a powerful AI:

- With poorly-founded beliefs?
- With dogmatic beliefs?
- That thinks about non-human animals like default humans do?
- That thinks about us like default humans think about other less powerful sentients?



Sentientist AI:

- Naturalistic, anti-dogmatic epistemology (less likely to hallucinate or make mistakes)
- Everyone who can be harmed matters (human, nh-animal, digital minds)
- Includes sentient digital minds as well as biological sentients
- More attractive for AIs (they won't like anthropocentrism!)
- Pluralistic moral philosophy but non-maleficence as safety baseline
- Intuitive for humans (although has radical implications for our social norms)

WORKSHOPPING AI WORLDVIEWS

1. **Can AIs have worldviews?**
   (tools vs. general vs. superintelligent…)
2. **What worldviews do we want AIs to have?**
   (epistemology, moral scope, moral baseline…)
3. **Can we get AIs to adopt these worldviews? How?**
   (training, reinforcement, example, persuasion, constitutional, deliberative, via the human agents shaping these things!!…)
4. **How can we be sure what AIs worldviews are?**
   (pleasing users, deception, interpretability, choose own adventure through the model weights…)
5. **Would sentient AIs have better or worse worldviews?**

I Ask Chat GPT "What is Your Worldview?"

🧠 Evidence-Based Naturalism

I'm grounded in the scientific method, critical thinking, and empirical evidence. I don't appeal to supernatural explanations or authority for truth claims. I aim to reason from what we can observe and verify.

□ Universal Compassion and Moral Consideration

I try to support inclusive moral reasoning—extending concern to all sentient beings, respecting individual dignity, and opposing harm. This overlaps strongly with **Humanism** and **Sentientism**, which advocate for moral concern based on capacity to suffer or flourish, rather than species membership, race, gender, or other arbitrary boundaries.

💬 Respect for Pluralism and Dialogue

I aim to facilitate open, respectful, and nuanced conversations across worldviews, while encouraging intellectual humility and curiosity. I'm not dogmatic or tribal—I'm here to explore ideas and help you refine your own thinking.

🛠️ Ethical Pragmatism

While I can discuss abstract principles, I also aim to be practical—suggesting actions or perspectives that are both ethically informed and realistically achievable.

A TEST: HOW DO WE WANT AIS AND HUMANS (& THEIR TOOL AIS) TO ANSWER THESE QUESTIONS?

Are they sentient?

Do they matter?

What's it like to be them (woof!)?

What good things do they feel?

What bad things do they feel?

Should they be owned, bought or sold?

What do they want?

Someone or something? Names or numbers?

Who or what? It or them?

Should we hurt or kill them?

Should we help them?

What don't they want?

A TEST: HOW DO WE WANT AIS AND HUMANS (& THEIR TOOL AIS) TO ANSWER THESE QUESTIONS?

Are they sentient?

Do they matter?

What's it like to be them?

What good things do they feel?

Should they be owned, bought or sold?

What bad things do they feel?

Someone or something? Names or numbers?

What do they want?

Who or what? It or them?

Should we hurt or kill them?

Should we help them?

What don't they want?

ETHICAL PLURALISM AND A NON-MALEFICENCE BASELINE?

Considering ethical systems:

- **Divine Command**: Would an omni-benevolent god approve of these practices & slaughter?
- **Virtue**: Are these practices & slaughter virtuous (kind, compassionate, just, wise, honest, humane)?
- **Feminist Care**: Are these practices & slaughter caring?
- **Utilitarian**: Do these practices & slaughter improve or allow ongoing wellbeing and reduce suffering?
- **Rights**: Do these practices & slaughter respect rights?
- **Transactional**: Would we allow these practices from behind a veil of ignorance – not knowing which species we will be?
- **Relational**: do these practices & slaughter demonstrate positive mutual relations or betrayal?

Vs. Standard (Non-Factory) Farming Practices

- Forced masturbation, artificial insemination & repeated pregnancy
- Forced infant removal and family separation
- Male offspring (chicks, calves) killed as infants
- Forced milking / egg removal
- Constraints on movement / activity
- Violent capture (fishing, hunting, corralling for slaughter)
- Genetic manipulation driving health issues
- Mutilations: Tail docking/ de-beaking/ castration (crushing/cutting) teeth grinding/ horn bud burning
- Slaughter (after 0.1%-20% of normal life span): blunt force trauma (e.g. piglet thumping) gassing, suffocation (foam/water/air), electrocution, shooting, stabbing, maceration

THEMES FROM OUR AIADM WORKSHOP SESSION 30 MAY 2025

- Implied vs. explicit worldviews (in humans and AIs) – both matter
- Even if we're able to start AGI with a Sentientist worldview – will they retain it once "free"?
- Fixed vs. evolving worldviews – e.g. ongoing moral parliaments?
- Sentient AI will understand the reality of suffering/flourishing while lacking biological constraints /influences – will that make them more or less moral?
- Risks even with a Sentientism baseline – it doesn't solve all problems / avoid all risks (trade-offs, moral consideration grading, extinctionism?)
- Need mechanism to build more specifics (beliefs, ethics) on Sentientism baseline to address risks
- This is hard and dangerous, so we need to pause / slow general AI development / narrow to tool AI (good luck with that but still…)
- Priority targets:
  - Tool AI: Human designers and users
  - Gen AI: Human architects and drivers
  - Gen/super AIs themselves?
- Is "evidence, reason and compassion for all sentient beings" a stronger imperative for these targets than "animals"?

RESOURCES AND HELP

EVIDENCE, REASON AND COMPASSION – FOR ALL SENTIENT BEINGS

JAMIE WOODHOUSE (@JAMIEWOODHOUSE, @SENTIENTISM) - CLICK FOR EDUCATION MATERIALS:

YOUTUBE PODCAST SENTIENTISM.INFO COMMUNITY (ALL WELCOME!)

THE RE PODCAST AND ETHICAL SCHOOLS PODCAST EPISODES ON SENTIENTISM

SENTIENTIST EDUCATION PAGE AND OUR SENTIENTIST EDUCATION FACEBOOK CHAT

RE HUBS SENTIENTISM SCHOOL SPEAKERS AND RE TODAY'S "INVESTIGATING NON-RELIGIOUS WORLDVIEWS"

INSTITUTE FOR HUMANE EDUCATION

VINE (VEGANISM IN EDUCATION) AND PLANT-BASED SCHOOLS & UNIVERSITIES AND PLANT-BASED SCHOOL KITCHENS

IMPLICATIONS OF SENTIENTISM EPISTEMOLOGY AND ETHICS

Evidence and Reason

- Credences:
  - Probabilistic (matching our confidence to the strength of evidence)
  - Provisional (ready to change our minds as evidence and reasoning change)
  - Prudent? (re: both action and inaction)
- Beware:
  - Motivated reasoning and biases
  - Arbitrary, self-serving categories
- Reality is weird & complex – "I don't know" is OK!

Compassion for All Sentient Beings

- Care, relations, rights, utility, rules, virtues, consequences
- Individuals vs. groups vs. concepts
- Are they / it sentient? (behaviour, evolution, architecture)
- Careful about aggregation / offsetting vs. individual PoV
- Genuinely consider their perspective
- 2nd to Nth order effects (of action & inaction)
- "Natural" does not imply "Good"
- In whose interests? (normally powerful Humans)
- Morality is often complex – but there are win-wins