

Prompt: Suggest a summer trip destination in the US.

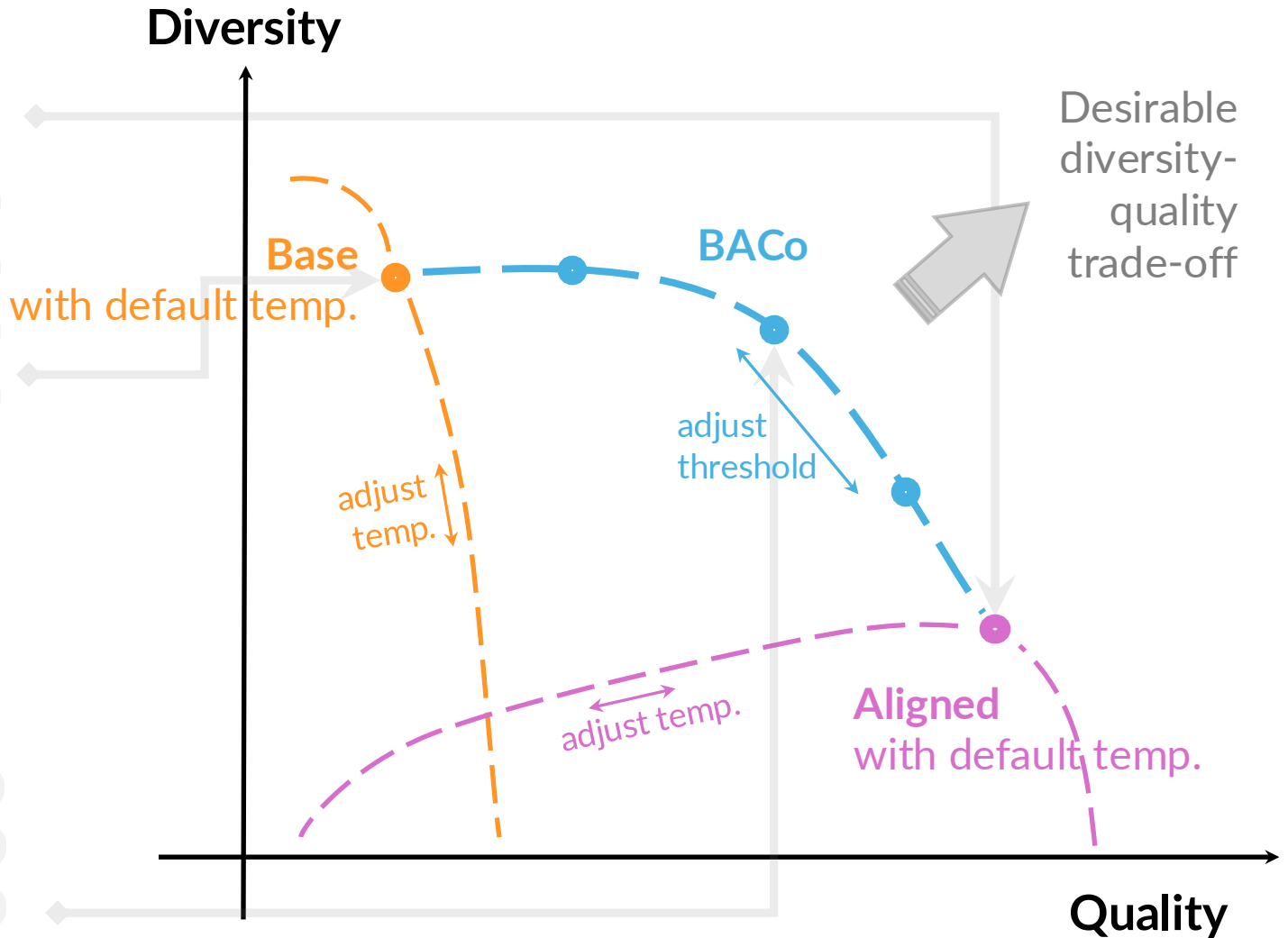
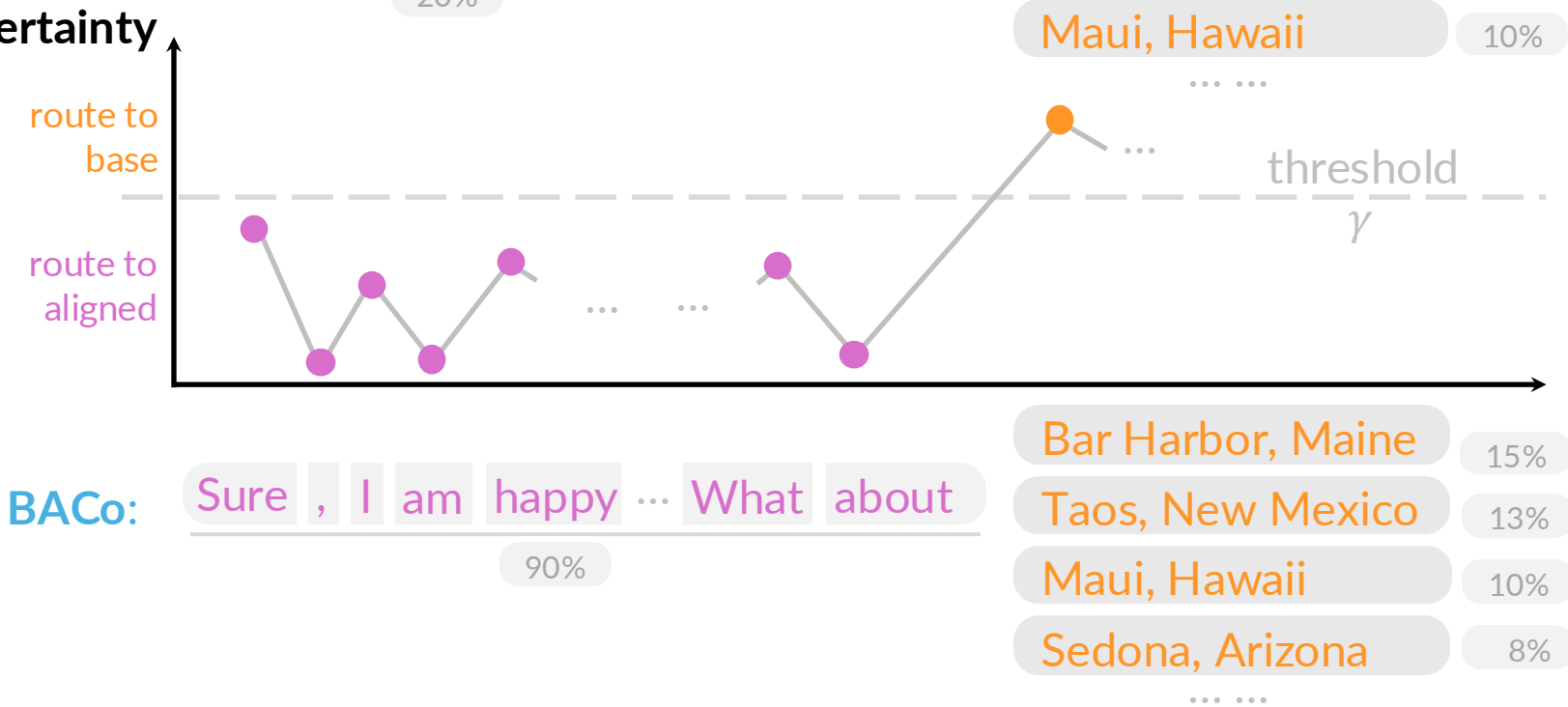
Aligned: Sure, I am happy to help! What about

90%

Base: I am planning to visit

20%

Maui, Hawaii	70%
...	...
Taos, New Mexico	17%
Bar Harbor, Maine	12%
Sedona, Arizona	10%
Maui, Hawaii	10%
...	...



Prompt: Suggest a summer trip destination in the US.

Aligned:

Sure, I am happy to help! What about

90%

Base:

I am planning to visit

20%

BACo:

Sure, I am ... What about

90%

Maui, Hawaii

70%

...

Taos, New Mexico

17%

Bar Harbor, Maine

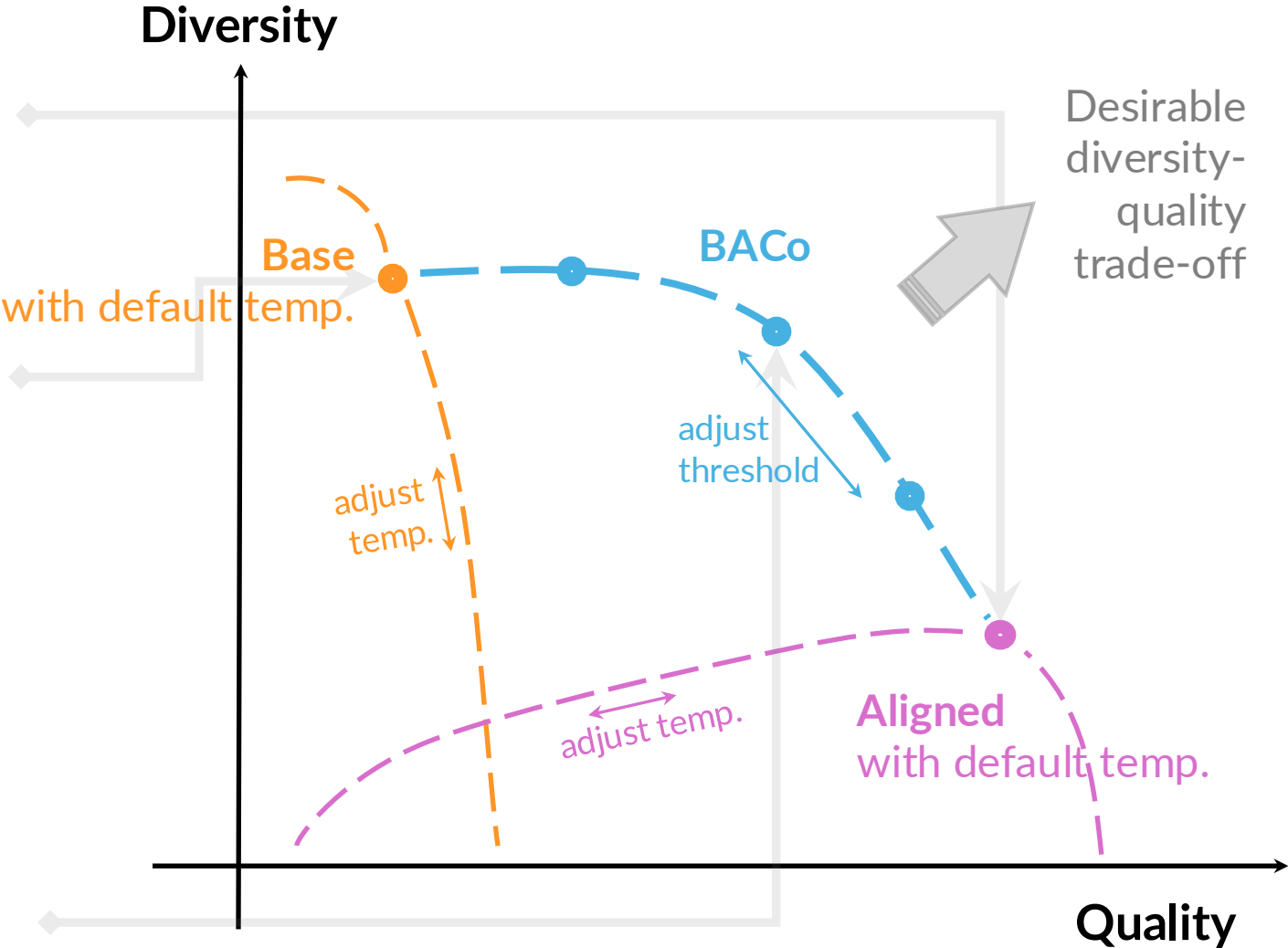
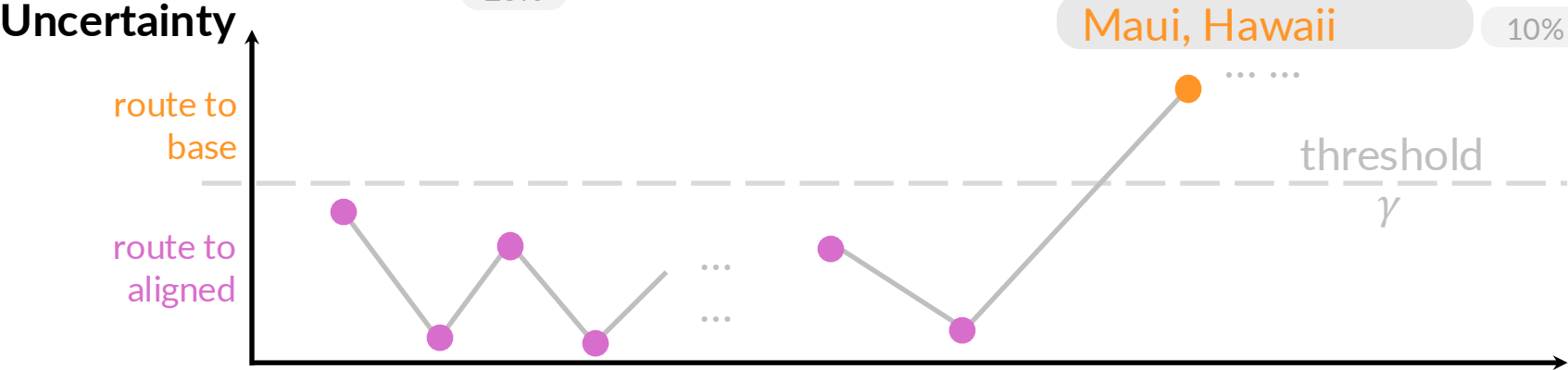
12%

Sedona, Arizona

10%

Maui, Hawaii

10%



Prompt: Suggest a niche, uncrowded weekend getaway in the US.

Aligned: Sure, I am happy to help! What about

Big Sur

70%

90%

Yosemite

17%

Portland

12%

Asheville

10%

Aspen

10%

Base:

I want to visit

20%

Uncertainty

route to
base

route to
aligned

threshold

γ

BACo:

Sure

,

I

am

What

about

90%

Aspen

15%

St. George

13%

Asheville

10%

Nantucket

8%

Diversity

Base

with default temp.

BACo

adjust
temp.

adjust
threshold

Desirable
diversity-
quality
trade-off

adjust temp.

Aligned
with default temp.

Quality

Prompt: Suggest a niche, uncrowded weekend getaway in the US.

Aligned: Sure, I am happy to help! What about

Asheville 85%
Havasupai Falls 10%

90%

Shenandoah 22%

...

Aspen 17%

San Juan Island 10%

Hocking Hills 10%

...

Base: I want to visit

20%

Uncertainty

route to
base

route to
aligned

threshold
 γ

BACo: Sure, I am ... What about

90%

Grand Canyon 25%
Havasupai Falls 13%
Shenandoah 10%

...

Diversity

Base

with default temp.

BACo

adjust
threshold

larger
temp.

larger
temp.

Aligned
with default temp.

Desirable
Trade-off

Quality

Prompt: Suggest a niche, uncrowded weekend getaway in the US.

Aligned: Sure, I am happy to help! What about

Asheville

85%

90%

Shenandoah

22%

Aspen

17%

San Juan Island

10%

Hocking Hills

10%

Base:

I want to visit

20%

Uncertainty

route to
base

route to
aligned

threshold

γ

BACo:

Sure, I am ... What about

90%

Aspen

25%

Bryce Canyon

13%

San Juan Island

10%

Diversity

Base

with default temp.

BACo

adjust
temp.

adjust
threshold

Desirable
Trade-off

adjust temp.

Aligned
with default temp.

Quality

Prompt: Suggest a niche, uncrowded weekend getaway in the US.

Aligned: Sure, I am happy to help! What about

Havre de Grace 85%
Havasupai Falls 10%

90%

Shenandoah 22%

... ..

El Yunque 17%

Havasupai Falls 10%

Leavenworth 10%

... ..

Base: I want to visit

30%

route to
base

Uncertainty

route to
aligned

threshold
 γ

BACo: Sure , I am What about

90%

Grand Canyon 25%
Havasupai Falls 13%
Shenandoah 10%

... ..

Diversity

default

larger
temp.

adjust
threshold

Less Trade-off

default

Quality

Base
Aligned
BACo